

## THESIS / THÈSE

### MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE À FINALITÉ APPROFONDIE

#### Caractérisation et analyse de la fonction du transcrit APOBEC3B Antisens1

Poncelet, Olivier

*Award date:*  
2020

*Awarding institution:*  
Université de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**Faculté des Sciences**

**CARACTERISATION ET ANALYSE DE LA FONCTION DU TRANSCRIT APOBEC3B-  
ANTISENS 1**

**Mémoire présenté pour l'obtention  
du grade académique de master 120 en biochimie et biologie moléculaire et cellulaire**

Olivier PONCELET

Août 2020



## **CARACTERISATION ET ANALYSE DE LA FONCTION DU TRANSCRIT APOBEC3B- ANTISENS 1**

PONCELET Olivier

### Résumé

Les APOBEC3 font partie d'une famille de cytidines désaminases et sont impliqués dans la réponse immunitaire innée contre de nombreux virus et rétroéléments. La perte de régulation de leur expression et en particulier celle de APOBEC3B est impliquée dans la cancérogenèse. Un ARN long non codant (ARNlnc), antisens au gène APOBEC3B appelé APOBEC3B-AS1 (A3B-AS1) a été découvert par annotation de données de RNA-seq. Cet ARNlnc a la particularité d'être le 191<sup>e</sup> transcrit le plus exprimé au stade E7 de l'embryogénèse. Son expression disparaît cependant très rapidement au cours du développement et il n'est retrouvé dans aucun tissu sain chez l'adulte. Il a déjà été montré que des longs ARN non codants antisens peuvent réguler l'expression d'autres gènes via divers mécanismes. Au cours de cette étude, nous caractérisons pour la première fois cet ARNlnc A3B-AS1 et nous analysons son impact sur la régulation de l'expression des gènes APOBEC3. Nous avons confirmé, par RT-PCR, l'existence de A3B-AS1 dans la lignée cellulaire lymphoïde cancéreuse KM-H2 ainsi que dans les iPSC. Nous avons cartographié avec précision ses extrémités par analyse 5' et 3' RACE. Des analyses d'un potentiel effet de A3B-AS1 sur les APOBEC3 ont été réalisées par surexpression et sous-expression de l'antisens. Il a également été observé qu'un traitement avec l'interféron- $\alpha$  ne semble pas modifier l'expression de A3B-AS1.

Mémoire de master 120 en biochimie et biologie moléculaire et cellulaire

Août 2020

**Promoteur:** N. Gillet



# Remerciements

---

Je voudrais dans un premier temps remercier mon promoteur, Nicolas Gillet, ainsi que mon superviseur, Florian Poulain, pour leur patience, leur disponibilité et ainsi que pour leurs nombreux conseils qui m'ont permis de progresser.

Je remercie également les autres membres de l'équipe APOBEC Noémie Lejeune et Kévin Willemart pour leurs conseils et leur aide durant ce mémoire.

Je souhaite remercier Damien Coupeau pour ses conseils avisés au laboratoire.

Je tiens également à remercier chacun des membres de l'URVI avec qui j'ai pu partager énormément de bons moments.

Je n'oublie pas non plus de remercier mes collègues mémorants Alexis Chasseur, Alexandra Decloux et Gabrielle Trozzi qui sont devenus de vrais amis durant ce mémoire.

Enfin, je souhaite remercier ma famille et mes amis pour leurs encouragements et leur soutien.



# Table des matières

---

Introduction .....	1
APOBEC3 .....	1
Généralités.....	1
APOBEC3B-AS1 .....	3
Rôle antivirale des APOBEC3 .....	3
Rôle dans la régulation des rétroéléments.....	5
Implication des APOBEC3 dans le cancer.....	7
Les longs ARN non codants.....	7
Généralités.....	7
Exemples de fonctions des ARNlnc.....	11
Objectifs .....	15
Matériel et Méthodes.....	19
Culture cellulaire .....	19
Traitement à l'interféron- $\alpha$ .....	19
Nucléofection des KM-H2 .....	19
Lipofection des HEK293T et mesure de la luminescence .....	19
Transduction des KM-H2 à l'aide de lentivirus .....	21
Synthèses des ADNc .....	21
Amplification de l'ADN par PCR.....	21
qPCR .....	21
RACE-PCR .....	23
Préparation d'ARN pour la 5' RACE-PCR et rétrotranscription .....	23
Rétrotranscription 3' RACE .....	23
Amplification par PCR nichée et séquençage des ADNc .....	23
Construction des plasmides.....	25
Plasmide pcDNA3.1 (-) A3B-AS1.....	25
Extractions protéiques et quantifications .....	25
Western Blot.....	25
Résultats .....	27
Caractérisation de A3B-AS1 .....	27
Mesure de l'expression de A3B-AS1 .....	27
Caractérisation des extrémités 5' et 3' d'A3B-AS1.....	28
Développement de KM-H2 surexprimant ou sous-exprimant le gène A3B-AS1 .....	31
Expression de shRNA contre A3B-AS1 dans les KM-H2.....	31





Surexpression de A3B-AS1 dans les KM-H2 .....	33
Étude de la fonction de A3B-AS1 .....	35
Caractérisation des isoformes de A3B dans les IPSC et les KM-H2 .....	35
Impacte de l'expression de A3B-AS1 sur la traduction de A3B, A3A et A3G dans les HEK293T .....	37
Impacte de l'interféron $\alpha$ sur l'expression de A3B-AS1 et A3B dans les KM-H2 .....	39
Discussion .....	41
Conclusion et perspectives .....	45
Bibliographie .....	51
Annexes .....	57



# Introduction

---

## APOBEC3

### Généralités

Les APOBEC3 proviennent d'une large famille de cytidines désaminases nommée APOBEC (apolipoprotein B mRNA catalytic polypeptide-like). Chez l'homme, cette famille comprend AID (activation induced deaminase), APOBEC1, APOBEC2, les 7 APOBEC3 (APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3DE, APOBEC3F, APOBEC3G et APOBEC3H) et APOBEC4. Il est intéressant de noter que la sous-famille des APOBEC3 est exclusivement retrouvée chez les mammifères.

Au cours de l'évolution, de nombreux gènes antiviraux dont la famille des APOBEC fait partie ont subi une sélection au contact de pathogènes viraux tel que les rétrovirus exogènes et rétro-éléments endogènes. Les conséquences de cette sélection sont l'existence de polymorphismes dans les régions codantes de ces gènes ainsi que la diversité des APOBEC [1], [2].

Certains APOBEC (A1, AID, A3C et A3H) possèdent un domaine cytidine désaminase zinc dépendante tandis que les autres en possèdent 2 (A3B, A3DE, A3F et A3G). Ils ont donc la capacité de changer les résidus cytidine en uracile au niveau des séquences d'ADN ou d'ARN simple brin.

Bien que les APOBEC possèdent une structure homologue, leurs fonctions sont très variées. Par exemple, AID est impliqué dans le phénomène d'hypermutation somatique. Il induit des mutations dans les lymphocytes B au niveau des régions V(D)J produisant les immunoglobulines. Ces mutations ont pour effet d'accroître la diversité des anticorps et donc de renforcer le système immunitaire adaptatif [3].

Chez l'homme, les APOBEC3 sont réunis sur un même locus au niveau du chromosome 22. Ces gènes sont ainsi issus de la duplication et de la modification d'un gène ancestrale. Une fois exprimées, ces enzymes ont différentes localisations subcellulaires : A3G, A3F et A3DE sont cytoplasmiques ; A3A, A3C et A3H peuvent être trouvés soit dans le cytoplasme soit dans le noyau, et A3B est spécifiquement localisé dans le noyau. Les APOBEC3 sont principalement connus pour être impliqué dans la défense contre certains virus exogènes ainsi que certains rétro-éléments endogènes. L'APOBEC3B est également étudié pour son implication dans le développement de nombreux cancer [4]. Des différences existent entre les cibles des différents APOBEC3. Alors que A3G mute préférentiellement les résidus cytidines dans les motifs CC, les autres APOBEC ciblent les cytidines dans les motifs CT. Les APOBEC3 ciblent uniquement l'ADN à l'exception de A3A qui crée également des mutations dans l'ARN [5].

Alors que les gènes APOBEC3 sont exprimés de façon constitutive dans la plupart des tissus, leur surexpression peut être induite par différents mécanismes de manière spécifique au type cellulaire. Par exemple, les gènes A3A, A3B et A3G peuvent être activés par des interférons, des virus à ADN, des infections, des dommages dans l'ADN, du LPS, différentes interleukines,... mais on sait encore peu de choses sur les voies de signalisation conduisant à la régulation de l'expression des gènes APOBEC3 [6]. Ces enzymes sont majoritairement



exprimées dans les populations de cellules immunitaires. Par exemple, la protéine antivirale dominante dans les lymphocytes T CD4<sup>+</sup> est APOBEC3G [7]. L'interféron  $\alpha$  induit la surexpression de A3G, A3F et A3DE dans les macrophages. APOBEC3A est exprimé dans les monocytes alors qu'il est absent dans les macrophages différenciés [8].

## **APOBEC3B-AS1**

Au niveau du locus APOBEC3, en position antisens du gène APOBEC3B, un ARN long non codant (ARNlnc) nommé APOBEC3B antisens 1 a été identifié par analyse systématique de RNAseq et annoté A3B-AS1. Cet ARNlnc A3B-AS1 et l'ARNm A3B montrent une complémentarité queue à queue. En raison de la similitude entre les différents gènes APOBEC3s, l'ARNm de A3B-AS1 est également partiellement complémentaire à A3A et A3G. L'annotation de A3B-AS1 dévoile qu'il est composé de 5 exons couvrant 702 paires de bases. L'ARN A3B-AS1 semble absent dans les tissus adultes sains. Sur une liste de 934 lignées de cellules cancéreuses, cet ARNlnc a été détecté par RNA-seq dans seulement 2 lignées (DND-41 et KM-H2). L'expression de A3B-AS1 a également été détectée dans la lignée embryonnaire de cellules souches (H1-hESC) et dans la lignée de cellules souches pluripotentes induite (IPSC). Des analyses de RNA-seq single cell des embryons humains en phase préimplantatoire ont également montré que cet ARNlnc est très spécifiquement exprimé au cours du stade E7 de l'embryogenèse. En effet, sur 26500 gènes analysés, A3B-AS1 est le 191<sup>e</sup> transcrit le plus exprimée dans l'épiblaste (également appelé ectoderme primitif) au stade E7 alors qu'il est le 910<sup>e</sup> au stade E6 et 6025<sup>e</sup> au stade E5. Ce gène est également exprimé dans l'hypoblaste (endoderme primitif) et le trophoctoderme, mais moins significative [9]. Il a déjà été montré que des longs ARN non codant antisens peuvent réguler l'expression d'autres gènes via divers mécanismes [10], [11].

## **Rôle antivirale des APOBEC3**

La famille des APOBEC3 a été décrite comme intervenant dans la réponse immunitaire innée contre différentes familles de virus telles que les lentivirus, les parvovirus, les herpesvirus ou encore les papillomavirus [8].

A3G est un acteur important dans la réponse à l'infection par le VIH. Lors de la formation des virions, A3G est empaqueté dans la capsid virale et inhibe l'infection par mutation de l'ADN complémentaire simple brin en cours de retro-transcription du génome viral [12]. A3D, A3F et A3H pourraient également être impliqués dans la réponse contre le VIH. Par ailleurs, ce virus possède un facteur nommé Vif (viral infectivity factor) capable de cibler ces APOBEC3 [13]. Ce facteur provoque l'ubiquitinylation des APOBEC3 et leur dégradation par le protéasome. Lorsque le gène vif est déficient, les APOBEC3 sont empaquetés dans les virions. A3H possède différents haplotypes (I-IV) au sein de la population. L'expression de l'haplotype II de A3H est corrélée avec une plus forte résistance au VIH [14].

Des mutations léthales provoqueront la dégradation des protéines virales tronquées par le système protéolytique de la cellule. Des mutations non léthales peuvent cependant conduire à la formation de nouveaux gènes viraux [15].



En plus d'être capable de provoquer des mutations dans les séquences d'ADN ou d'ARN de ces virus, les APOBEC3 peuvent être impliqués dans l'immunité innée de manière indépendante de leur activité cytidine déaminase. Certaines études montrent que les APOBEC3 pourraient inhiber l'élongation et l'accumulation de produit de retro-transcription du VIH et du MLV (murin leukemia virus) dans le cytoplasme [16], [17].

L'expression des APOBEC3 peut également être induite par une réponse aux interférons. Les cellules dendritiques et les macrophages ont pour rôle de détecter l'ADN étranger non méthylé par l'intermédiaire du TLR9 (toll-like receptor 9) présent dans les endosomes. D'autres senseurs de l'ADN sont impliqués tels que AIM2 (absent in melanoma-2), STING (stimulator of IFN genes), DAI (DNA-dependent activator of IFN regulatory factor) et l'ARN polymérase III. Ensemble, ils participent aux voies de signalisation qui produisent des cytokines pro-inflammatoires telles que et les interférons  $\alpha$  ce qui entraîne la transcription de gènes effecteurs afin de générer une large réponse immunitaire innée [8]. A3B, A3F et A3G sont connus pour être induit par ce mécanisme dans les hépatocytes lors d'infection par HBV (hepatitis B virus). A3A peut également être induit dans les monocytes par réponse interféron et pourrait être un inhibiteur du parvovirus AAV-2 (adeno-associated virus type 2) [5].

## **Rôle dans la régulation des rétroéléments**

Les rétrotransposons constituent 40% environ du génome humain. Ceux-ci sont répartis en 2 groupes principaux : les rétrotransposons à LTR (Long Terminal Repeat) tels que les retrovirus endogènes (ERV) et ceux ne possédant pas de LTR tel que les LINEs (Long interspaced nuclear elements) et SINEs (Short interspaced nuclear elements)[18].

Dans les tissus somatiques, les régions du génome contenant des rétrotransposons possèdent des modifications épigénétiques telles que la méthylation de l'ADN et la méthylation histone permettant d'inhiber leur expression. Lors des premières étapes de l'embryogénèse, aux environs du stade 8 cellules, ces régions sont réactivées. En effet, lors de cette phase, les histones de nombreuses régions du génome sont déméthylées permettant un décompactage de la chromatine et la liaison des facteurs de transcription ainsi que du complexe de la polymérase [19].

Bien que leurs fonctions n'aient pas encore été définies, de plus en plus d'études montrent que les rétrotransposons pourraient avoir un rôle dans différents processus biologiques. Ils peuvent interférer avec l'expression des gènes par leur transcription en antisens ou en provoquant une terminaison prématurée de la transcription. Ils peuvent également fournir de nouveaux sites d'initiation de la transcription ou des éléments régulateurs tels que des régions activatrices (enhancers) et ainsi modifier la régulation et la structure des gènes. Les rétrotransposons sont également capable de médier la recombinaison homologue[20].

Les HERVKs (rétrovirus endogènes humains K) sont transcrits durant l'embryogénèse humaine. Leur expression débute lors de l'activation du génome embryonnaire au stade de huit cellules, continue pendant l'émergence des cellules épiblastiques dans les blastocystes en phase préimplantatoires et cesse lors de la différenciation des cellules souches embryonnaires au niveau du blastocyste [21]. Ceci correspond à l'augmentation progressive de l'expression de





A3B-AS1 (selon les données de séquençage d'ARN de NCBI). Fait intéressant, à partir du navigateur de génome UCSC, il a été observé qu'une séquence de HERVK LTR est située directement en amont du premier exon de A3B-AS1. En outre, des données de Chip-seq ont prédit que le facteur de transcription SP1 permettrait la transcription de ces LTR dans la lignée H1-hESC.

Il est tentant de spéculer que c'est le même mécanisme qui amène à la transcription de HERVK et A3B-AS1 pendant les premières étapes de l'embryogenèse humaine. En effet, pendant l'embryogenèse, le facteur de transcription SP1 peut se lier aux promoteurs des rétroéléments pour les activer. Dans un même temps, A3B inhibe la rétrotransposition. SP1 se lierait au promoteur de A3B-AS1 pour en activer la transcription. Ce mécanisme pourrait avoir pour objectif de permettre l'inhibition de l'expression de A3B pendant la réactivation des rétrotransposons par son antisens.

## **Implication des APOBEC3 dans le cancer**

Dû à leur fonction cytidine désaminase, ces enzymes sont également capables de muter l'ADN génomique lorsqu'il se trouve sous forme simple brin. Des analyses par NGS (next-generation sequencing) ont permis de montrer que plus de 50% des cancers possèdent une signature APOBEC3 .. Il a été rapporté que les tumeurs provenant de cancers du poumon et surexprimant A3B possèdent en moyenne 2 fois plus de mutations que celles exprimant faiblement A3B. Les mutations dans l'ADN des cellules tumorales ont pour conséquence de provoquer le développement de résistances thérapeutiques [22] Par exemple, une réduction de l'efficacité des traitements au tamoxifène a été mise en corrélation avec la surexpression de A3B [23], [24].

## **Les longs ARN non codants**

### **Généralités**

Les longs ARN non codants représentent un vaste groupe d'ARN possédant une taille supérieure à 200 nucléotides et ne codant pas pour une protéine. Les ARN messagers (ARNm) ne représentent que 36% des ARN totaux. Parmi les ARN non codants, les longs ARN non codants (ARNlnc) sont les plus fréquemment retrouvés (60%). En comparaison, les miARN ne représentent que 19% de ce groupe[25].

L'ensemble des ARN non codants sont très étudiés pour leurs capacités de régulation dans l'expression de nombreux gènes. Le projet ENCODE (ENCyclopedia of DNA Elements) a notamment permis de faire un recensement de toutes les séquences codantes du génome humain. De cette étude, il est ressorti que 90% du génome est transcrit mais seulement 2% est traduit [26].

Les ARN non codants ont été par la suite classifiés en sous-types. Parmi ceux-ci sont retrouvés l'ARN ribosomal (ARNr), l'ARN de transfert (ARNt), les microARN (miARN), les longs ARN non codants (ARNlnc), les ARN interagissant avec PIWI, les ARN circulaires et les petits ARN nucléolaires. Pour la plupart, ces ARN non codants auraient pour origine la duplication et la



modification de séquences codantes ou d'éléments transposables. Les ARNlnc ont une stabilité comparable à celle des ARNm. Leur localisation cellulaire est généralement nucléaire bien que certains peuvent être trouvés dans le cytoplasme. Leur structure secondaire est très conservée et souvent complexe. Par exemple, le ARNlnc HOTAIR retrouvé uniquement chez les mammifères ne partage que 58% d'homologie entre l'homme et la souris mais possède une structure similaire entre les deux espèces. L'importance de la conservation de la structure de HOTAIR et non de sa séquence d'ADN est liée à la conservation des sites de liaison protéines - ARNlnc [27].

Pendant longtemps, les ARNlnc étaient considérés comme étant du bruit de fond dans les résultats de séquençages d'ARN. Cette idée provenait en partie du fait qu'ils sont souvent moins exprimés que les ARN codants. Cela dit, il a par la suite été remarqué que leur expression était parfois transitoire que ce soit pour réguler certaines étapes du développement ou dans certains mécanismes cellulaires. En ajoutant à cela que leur expression est souvent spécifique au type cellulaire, il a été dévoilé que les premières études de transcriptomique sous-estimaient leur expression. En effet, ces études se faisaient à partir d'une population cellulaire hétérogène diluant ainsi de potentiels hauts taux d'expression retrouvés dans une faible quantité de cellules. Ce n'est qu'avec l'arrivée de méthodes tel que le *single cell RNA sequencing* que des études plus poussées sur l'expression des ARNlnc ont pu être mises en oeuvre [28].

Au vu de l'étendue du groupe des ARNlnc de nombreux critères peuvent être utilisés pour les classer : la taille, la localisation génomique, la localisation cellulaire, la fonction, le lien avec différentes voies métaboliques, ... Dans le cadre de cette étude sur A3B-AS1, le premier critère à analyser est sa localisation dans le génome. Tout d'abord, les ARNlnc peuvent être intergéniques ou intragéniques. Les ARNlnc intragéniques peuvent être séparés en différentes catégories : chevauchants, introniques et antisens. Les ARNlnc chevauchants partagent partiellement ou entièrement leur séquence de lecture avec un autre gène. Les ARNlnc introniques possèdent l'entièreté de leur cadre de lecture dans une séquence intronique d'un autre gène. Et enfin, les ARNlnc antisens tels que A3B-AS1 sont codés sur le brin opposé au gène sens. Ces ARN antisens peuvent avoir une fonction en cis dans le cas où ils régulent le gène dont ils sont antisens ou une fonction trans dans le cas où le ou les gène(s) qu'ils régulent proviennent d'une autre séquence génomique. Les ARN antisens peuvent également être head to head lorsque seul les régions 5' des séquences sont complémentaires, tail to tail lorsque seul les régions 3' des séquences sont complémentaires ou interne lorsque leur séquence est complètement recouverte par le transcrit sens. En reprenant ces différents critères, A3B-AS1 peut être décrit comme un long ARN non codant antisens tail to tail à A3B et possédant un promoteur indépendant. Celui-ci ne permet donc pas la transcription d'un autre gène sur le brin sens (promoteur bidirectionnel) et ne se situe pas dans la séquence d'un autre gène (promoteur cryptique). Les ARNlnc sont transcrits de manière similaire aux ARNm classiques. Pour la plupart, ils sont transcrits par l'ARN polymérase II et peuvent subir différentes modifications post-transcriptionnelles telles que l'ajout d'une coiffe en 5', d'une séquence polyadénylée en 3' ou encore de l'épissage normal ou alternatif [29].



## Exemples de fonctions des ARNlnc

Les ARNnc possèdent un rôle de régulation dans de nombreux processus biologiques. Ils ont été retrouvés chez tous les organismes vivants ainsi que chez les virus. De manière intéressante, une étude a démontré qu'il existe une corrélation entre la complexité des organismes eucaryotes et le nombre d'ARNnc qu'ils expriment. Cette corrélation n'est cependant pas retrouvée pour les ARN codants. Par exemple les nématodes possèdent tout comme les humains approximativement 20000 gènes codants mais moins de 50% de leur génome est transcrit en ARNnc [30]. Ces résultats suggèrent qu'une grande diversité d'ARNnc permet d'augmenter le nombre de voies de régulation de l'expression des gènes et serait donc un atout majeur dans l'évolution d'organismes complexes.

Les différentes caractéristiques des ARNlnc telles que leur orientation, leur stabilité, leur localisation subcellulaire ou encore leur structure influencent leurs multiples rôles dans la cellule. La plupart des ARNlnc dont la fonction a été décrite ont été retrouvés dans le noyau. Ceux-ci sont impliqués dans la régulation de processus tels que l'épigénétique, la transcription ou encore la post-transcription. Les ARNlnc cytoplasmiques sont quant à eux retrouvés en tant que régulateurs de la post transcription et de la traduction en modulant la stabilité des ARNm, en servant d'éponge à miARN, ou en liant certains complexes protéiques [31]. Au vu des nombreux mécanismes de régulation existants, seules les principales fonctions des ARN longs non codants antisens seront traitées ici.

Tout d'abord, ces transcrits peuvent influencer l'initiation de la transcription d'autres gènes. La transcription des ARNlnc peut provoquer une compétition entre son promoteur et celui d'une séquence codante proche. En effet, l'assemblage de la machinerie de transcription au niveau d'un promoteur peut créer une obstruction empêchant la liaison d'une seconde machinerie.

Certains peuvent causer la dégradation du transcrit du sens en formant un duplex d'ARN avec ce dernier qui est alors dégradé par les RNases III. D'autres ARNlnc peuvent se lier au transcrit du sens et masquer le site de liaison des microARN. C'est le cas pour l'antisens de BACE1 (BACE1-AS) qui inhibe la liaison de miR-485-5p à BACE1 [32].

Certains antisens sont connus pour provoquer des modifications épigénétiques par liaison avec le complexe polycomb PRC2. En effet, la structure secondaire et tertiaire de certains ARNlnc leur permet de se lier à certains complexes protéiques. Par exemple, l'antisens ANRIL (antisens de CDKN2B) recrute PRC2 en cis qui induit une méthylation des histones H3K27 dans les transcrits des gènes CDKN2B et CDKN2A connus pour être des suppresseurs de tumeurs. Un autre exemple de modification épigénétique par un antisens est la répression de l'homeobox HOXD en trans par le complexe PRC2 lié à HOTAIR (antisens de HOXC) [10].

Il a également été démontré que la transcription de certains antisens simultanément au gène sens pouvait créer une interférence par interférence des machineries transcriptionnelles. C'est le cas pour IME4, un gène codant pour un régulateur de la méiose chez la levure. Sa transcription peut être bloquée par la transcription de son antisens RME2. Ces 2 gènes ne possèdent que 450 paires de base complémentaires et l'antisens ne couvre pas la séquence promotrice du gène sens, suggérant que seul l'élongation est inhibée et non l'initiation [33].



Ces fonctions précédemment décrites pourraient chacune être des pistes de réflexions dans la recherche de fonction de A3B-AS1. De nombreux autres mécanismes d'actions des ARNinc antisens ont été découverts. Parmi ceux-ci se trouve l'épissage alternatif ou encore la régulation des séquences promotrices de la transcription ou de la traduction [10]. Les caractéristiques connues de notre séquence et de son locus nous permettent déjà d'écarter ces options.





# Objectifs

---

Les objectifs de ce mémoire sont la caractérisation et l'étude de la fonction de A3B-AS1. Dans un premier temps, nous avons choisi les lignées cellulaires qui serviront de modèle pour la suite de notre recherche. Les données de RNA-seq disponibles sur NCBI ont permis de déterminer que cet ARNnc est présent dans les cellules H1-hESC, les iPSC, les DND-41 ainsi que dans les KM-H2. Dans le cadre de ce mémoire, les H1-hESC n'ont pas été étudiées. En effet, l'établissement d'une culture de cette lignée s'avère plus longue et fastidieuse. Il a donc été décidé de se concentrer sur les autres lignées cellulaires. La première étape sera de confirmer par nous-même la présence de A3B-AS1 dans ces lignées. Pour ce faire, l'expression d'ARN de chaque lignée sera analysée par RT-PCR et quantifiée par RT-qPCR.

Une fois l'expression de A3B-AS1 confirmée dans nos lignées, nous souhaitons caractériser notre séquence en étudiant de potentielles isoformes de celle-ci ainsi qu'en analysant ses extrémités 5' et 3'. En effet, bien que A3B-AS1 aient été découvert par RNA-seq, aucune étude approfondie n'a encore été réalisée sur cet antisens. Grâce à la mise au point d'une RACE PCR, nous verrons que l'annotation de A3B-AS1 retrouvée sur NCBI se révèle être incorrecte et incomplète (voir résultats).

Après la caractérisation de A3B-AS1 et la sélection des lignées modèles à étudier, nous souhaitons déterminer la fonction de notre antisens. Pour cela, nous commencerons par rechercher une potentielle interaction entre celui-ci et A3B ainsi que les autres APOBEC avec lesquels il est partiellement complémentaire. Pour ce faire, nous analyserons l'expression de ces APOBEC dans nos lignées cellulaires modèles. Ces analyses se feront par RT-qPCR pour observer la présence ou non de transcrits et puis par Western-blot pour déterminer si les formes protéiques de ces apobec sont présentes dans nos lignées. Il nous sera également nécessaire de déterminer quel isoforme de A3B est présente dans nos lignées modèles. En effet, A3B est connu pour posséder 4 isoformes. Ces isoformes possèdent des séquences différentes au niveau de l'extrémité complémentaire à A3B-AS1.

Afin d'analyser la fonction d'A3B-AS1, nous allons le surexprimer et le sous-exprimer et analyser les effets de ces changements d'expressions sur ses cibles potentielles.

Concernant la sous-expression de l'antisens, nous avons décidé de construire et d'intégrer des shRNA dirigé contre celui-ci dans les KM-H2. Ces shRNA sont transduits à l'aide de Lentivirus précédemment produit dans des HEK293T.

Pour permettre d'analyser les effets de l'expression de A3B-AS1 sur A3A, A3B et A3G nous avons mis au point une méthode de surexpression transitoire de ces APOBEC3 lié à une luciférase ainsi que de l'antisens dans des HEK293T par lipofection. Cette lignée cellulaire a été choisie car elle n'exprime pas l'antisens naturellement. Avec cette méthode il est donc possible de mesurer le niveau d'expression des APOBEC3 à travers la mesure de l'activité de la luciférase avec ou sans la présence de A3B-AS1.

Nous allons également essayer de créer une lignée stable de KM-H2 surexprimant A3B-AS1 à l'aide d'un plasmide intégré par nucléofection. Avec cette lignée, nous espérons au moins décupler l'expression de notre antisens pour observer par la suite son impact sur les APOBEC3.



Enfin, nous souhaitons étudier l'effet de la présence de l'interféron-  $\alpha$  sur l'expression de A3B-AS1. En effet, il est connu qu'une surexpression de certains APOBEC peut être induite par une réponse aux interférons-  $\alpha$ . Il ne semble donc pas improbable que l'expression de notre antisens soit également impactée par une réponse cellulaire de ce type.

Ces expérimentations ont comme premier objectifs de caractériser de manière complète et détaillée A3B-AS1. Le second objectif est de déterminer quelles sont les lignées cellulaires adéquates pour l'étude de cet antisens. Enfin, le dernier objectif est de créer des données préliminaires pour fournir des pistes de fonction de A3B-AS1.



# Matériel et Méthodes

---

## Culture cellulaire

Les lignées cellulaires KMH2 (lignée cellulaire de carcinome thyroïdien humain anaplasique) et DND-41 (cellules T leucémiques) ont été cultivées dans du milieu Roswell Park Memorial Institute 1640 (RPMI-1640) auquel a été ajouté 10% de FBS et 1% de pénicilline/streptomycine. La lignée cellulaire HEK293T (human embryonic kidney) a été cultivée dans du Dulbecco's Modified Eagle Medium (DMEM) auquel a été ajouté 10% de FBS et 1% de pénicilline/streptomycine. La lignée cellulaire d'IPSC de la collection de l'Institut Allen (Aics0011) a été fournie par D Dutraux du laboratoire URBC de l'UNamur. La lignée cellulaire IPSC a été cultivée dans du milieu mTeSR<sup>TM</sup> Plus (stem cell<sup>TM</sup> technology).

## Traitement à l'interféron- $\alpha$

Les cellules KMH2 et DND-41 ont été traitées à l'interféron- $\alpha$  (1000/mL) lorsque la confluence cellulaire était de 70%. Les cellules traitées ainsi que des cellules contrôle non traitées ont été récoltées au temps 0, 1, 3, 6 et 24h. L'expression des gènes A3B-AS1, A3B et RSAD2 a été quantifiée par RT-qPCR et normalisée par rapport à la quantification de l'expression de GAPDH.

## Nucléofection des KM-H2

La surexpression de l'isoforme 3 de A3B-AS1 dans les KM-H2 a été faite par nucléofection du vecteur de surexpression pcDNA3.1 (-). Les KM-H2 utilisées pour la nucléofection sont préalablement mis en culture sans antibiotiques. La nucléofection est faite à l'aide du kit V cell line nucleofector<sup>TM</sup> de chez Lonza. Pour la sélection des cellules transfectées, les cellules sont traitées avec l'antibiotique G418. Les KM-H2 ont également été transfectées avec le plasmide Pmax-GFP comme contrôle de l'efficacité de transfection.

## Lipofection des HEK293T et mesure de la luminescence

Des HEK293T ont été transfectées à une confluence d'environ 70% par lipofection avec le kit Lipofectamine Transfection Reagent. Les vecteurs utilisés sont le psi 2 contenant les séquences complètes de A3B, A3A ou A3G et pcDNA3.1 (-) contenant la séquence de A3B-AS1. Le milieu de culture est changé 5 heures après la lipofection. Après 48 heures, les HEK293T sont lavées avec du PBS puis traitées avec une solution de PLB (Passive Lysis Buffer). Les cellules sont ensuite placées sur agitateur durant 15 minutes à 50RPM. Le lysat obtenu est centrifugé 30 secondes à 10000 g et le surnageant est récolté. La mesure de la luminescence de la *firefly* luciférase ainsi que de la *Renilla* luciférase est réalisée à l'aide du kit Dual luciferase reporter assay de Promega et du Luminoskan Ascent Thermo Scientific. Tout d'abord, 50 $\mu$ l de Luciferase Assay Reagent II sont ajoutés à 10 $\mu$ l des extraits protéiques produits et la luminescence de la *firefly* luciférase est mesurée. Ensuite 50 $\mu$ l de Stop & Glo Reagent sont ajoutés pour mesurer la luminescence de la *Renilla* luciférase.



## Transduction des KM-H2 à l'aide de lentivirus

La production de shRNA ciblant A3B-AS1 par les KM-H2 a été permise par infection avec des lentivirus. Des HEK293T préalablement mise en culture dans du RPMI1640 complémenté avec 10% de FBS et 1% de pénicilline/streptomycine sont transfectées par lipofection avec le kit Lipofectamine Transfection Reagent. Les vecteurs utilisés sont le pCMV-VSV-G, le psPAX2 ainsi que 3 exemplaires du pSicoRMS2 capable chacun de produire un shRNA ciblant A3B-AS1. Le milieu des HEK293T contenant les lentivirus est récolté 48 heures et 72 heures après la transfection. Les KM-H2 sont mise en culture avec ce milieu tout de suite après sa récolte. Les KM-H2 transduites sont sélectionnées en ajoutant de la blasticidine (10 µg/ml) à leur milieu. Une lignée de KM-H2 produisant des shRNA scramble est également produite et sert de contrôle négatif.

## Synthèses des ADNc

L'ARN total a été extrait avec le kit TRI Reagent<sup>®</sup> de Promega<sup>®</sup> puis traité 15 minutes à 37°C à la DNase. Un traitement au phénol chloroforme est ensuite exécuté pour purifier les échantillons. La SuperScript<sup>™</sup> III Reverse Transcriptase (Thermo Fisher) a été employée pour la rétrotranscription des ARN Poly(A). 1µg d'ARN total a été incubé avec l'enzyme et les oligos dT durant 50 minutes à 50°C. la réaction a été stoppée par inactivation thermique de l'enzyme pendant 5 minutes à 80°C.

## Amplification de l'ADN par PCR

L'expression cellulaire des ARNm d'A3B-AS1, A3B et GAPDH ont été analysées par PCR. Les séquences des amorces employés se trouvent dans la figure supplémentaire 1. Les ADNc ont été spécifiquement amplifiés en utilisant la GoTaq<sup>®</sup> DNA Polymerase (Promega). Les 35 cycles d'amplification PCR consistaient en 30 secondes à 98°C suivies de 30 secondes à la température d'hybridation des amorces puis 1 minute à 72°C et ont été suivis d'un cycle final d'élongation de 5 minutes à 72°C. Les produit PCR ont été visualisés par migration sur gel d'agarose 1%.

## qPCR

L'expression des ARNm d'A3A, A3B, A3B-AS1, A3G, RSAD2, GAPDH, HPRT et TBP (les séquences des amorces employés se trouvent dans la figure supplémentaire 1) a été mesurée à l'aide du mix FastStart DNA Master SYBR GREEN. et a été analysée par le système Eco<sup>™</sup> Real-Time d'Illumina<sup>®</sup>. L'amplification s'est déroulée comme suit : une dénaturation initiale de 5 minutes à 95°C suivie de 35 cycles de dénaturation (30s à 95°C), hybridation (30s à 55°C) et élongation (30s à 72°C). Les analyses ont été réalisées en duplicats à partir de trois niveaux de dilution du produit de retrotranscription (1/10, 1/30 et 1/90).





## RACE-PCR

### Préparation d'ARN pour la 5' RACE-PCR et rétrotranscription

Les extrémités 5' et 3' des longs ARN non codants d'A3B-AS1 ont été analysées par RACE-PCR en se servant du kit GeneRACER™ en suivant le protocole d'utilisation de la firme. Brièvement, la première étape a été de traiter 5µg d'ARN total à la CIP pendant 1h à 37°C pour éliminer les phosphates 5' des ARN non coiffés. Ensuite, les extraits d'ARN obtenus après purification au phénol chlorophorme et précipitation à l'éthanol ont été incubés 1h à 37°C avec l'enzyme TAP pour éliminer la coiffe d'ARN. Après une nouvelle étape de purification au phénol chlorophorme et une précipitation à l'éthanol, les ADNc ont été synthétisés par la rétrotranscriptase SuperScript™ III Reverse Transcriptase durant 50 minutes à 50°C avec un mélange de deux amorces spécifiques à A3B (amorces 5' RACE RT 1 et 2 localisées à la fin de l'exon 2). Enfin, l'enzyme a été inactivée par incubation à 80°C pendant 5 minutes.

### Rétrotranscription 3' RACE

5µg d'ARN total ont été rétrotranscrits dans les memes conditions que celles décrivent pour la 5'RACE. L'amorce utilisée pour la rétrotranscription est une amorce d'oligo dT GeneRACER™ qui correspond à 24dt nucléotides combinés à 36 bases spécifiques à la séquence de l'extrémité 5'.

### Amplification par PCR nichée et séquençage des ADNc

Les ADNc 5' et 3' RACE ont été amplifiés par PCR et PCR nichée successive. Pour chaque échantillon d'ADNc des 5' et 3' RACE, deux combinaisons différentes d'amorces PCR ont été utilisées pour la première PCR. Les détails des combinaisons d'amorces employées se trouvent dans le tableau supplémentaire 2. Les PCR et PCR nichées ont été réalisée comme reporté précédemment (PCR). Les produits d'amplification ont été purifiés par migration sur gel d'agarose 1%, avec le kit d'extraction Macherey-Nagel puis clonés dans le vecteur pGEM-T Easy de Promega. Les plasmides ont été électroporés dans *E. coli* et les bactéries ont été incubée overnight à 37°C dans du milieu LB-agar avec ampicilline (100µg/mL). Un screen PCR a été fait et les échantillons ont été envoyés à Eurofins pour séquençage.



# Construction des plasmides

## Plasmide pcDNA3.1 (-) A3B-AS1

La séquence complète de l'isoforme 3 de A3B-AS1 a tout d'abord été amplifiée par RT-PCR à partir d'une extraction d'ARN de KM-H2. Les amorces utilisées sont les amorces A3B-AS1 NheI FW et A3B-AS1 HindIII RV. La séquence obtenue possède le site de restriction NheI en 5' et le site de restriction HindIII en 3'. Les séquences obtenues sont purifiées à l'aide du kit NucleoSpin® Extract II de MACHEREY-NAGEL. La restriction de A3B-AS1 ainsi que du pcDNA3.1 (-) est faite à l'aide des enzymes de restriction NheI et HindIII de chez NEB. Le plasmide est déphosphorylé par traitement à la CIP et ensuite purifié par extraction phénol chloroforme et précipitation à l'éthanol. La ligation de l'insert et du plasmide est faite à l'aide de la T4 DNA Ligase de chez NEB. Les plasmides obtenus ont été électroporés dans *E. coli* et les bactéries ont été incubées overnight à 37°C dans du milieu LB-agar avec ampicilline (100µg/mL). Un screen PCR a été réalisé et les échantillons ont été envoyés à Eurofins pour séquençage. Après sélection du plasmide, une midiprep a été réalisée à l'aide du kit NucleoBond® Xtra Midi de MACHEREY-NAGEL pour amplifier celui-ci.

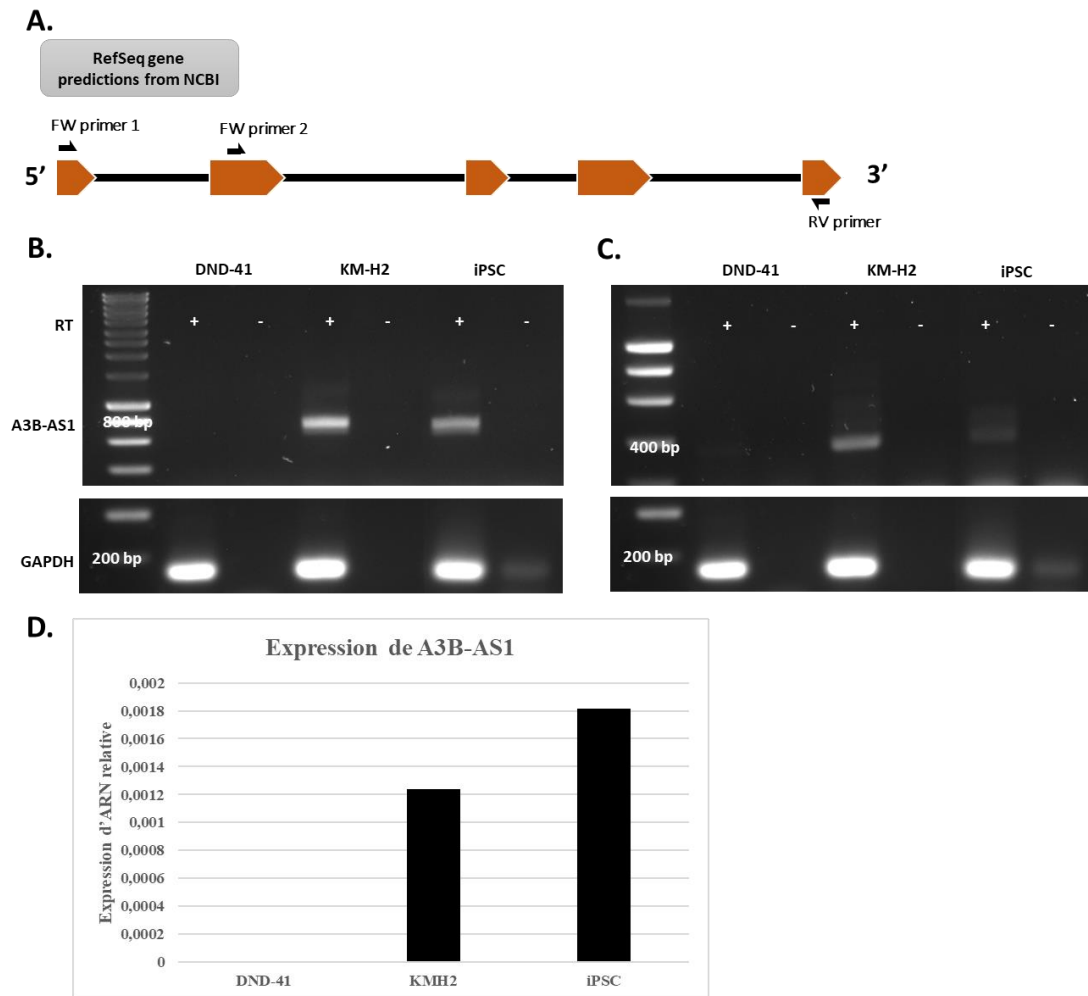
## Extractions protéiques et quantifications

Les protéines ont été extraites en utilisant le tampon RIPA supplémenté par un mélange d'inhibiteurs de protéases, du Triton X-100 10x et du PMSF 1mM 100x pour resuspendre les cellules. Après 30 minutes d'incubation sur glace, les protéines cellulaires ont été extraites par sonication durant 3 cycles de 30 secondes ON et 30 secondes OFF avec le dispositif Bioruptor® (Diagenode). Ensuite, les échantillons ont été centrifugés à 14 000 rpm à 4°C et le surnageant a été collecté et stocké à -80°C.

Les protéines ont été quantifiées à l'aide du kit Pierce™ BCA Protein Assay Kit (ThermoFischer) en suivant le protocole fourni par la firme. Les mesures ont été réalisées en utilisant le Multiskan EX (Thermo Electro Corporation) à 550nm.

## Western Blot

25µg d'extraits totaux de protéines ont été séparés sur gel d'acrylamide SDS-PAGE 10% pendant 2h à 120V. Les protéines ont été transférées à 4°C durant 1h10 à 200mA sur membrane de PVDF (PolyVinylDene Fluoride) précédemment activée avec du méthanol. La membrane a été bloquée avec du TBST-BSA 5% (Tris-Buffered Saline avec 0,1% de Tween-20) pendant 1h et incubée overnight à 4°C avec l'anticorps primaire dilué dans du TBST-BSA 5% (anticorps de lapin A3B et A3G de cell signaling concentré à 1/1000 et anticorps de lapin hsp-90 de Sigma concentré à 1/1000). Après 3 lavages de 10 minutes au TBST, la membrane a été incubée avec l'anticorps secondaire dilué dans du TBST-BSA 5% pendant 1h à température ambiante (immunoglobulines/HRP polyclonales de porc anti-lapin de Dakon concentré à 1/2000). Après 3 lavages de 15 minutes au TBST, la membrane a été incubée avec le substrat Pierce™ ECL Western Blotting Substrat (Thermo Scientific) pour hsp90 et le substrat SuperSignal™ West Femto Maximum Sensitivity Substrat (Thermo Scientific) pour A3B et A3G. la détection des protéines a été pratiquée par le ImageQuant LAS4000 mini (Ge Healthcare Life Sciences) avec caméra CCD refroidie.



**Figure 1: Analyse par RT-PCR et RT-qPCR de l'expression de A3B-AS1 sur les lignées cellulaires DND-41, KM-H2 et iPSC.** A. Représentation du locus A3B-AS1 et position des primers (exons en orange et introns en noir). B. RT-PCR réalisée avec l'amorce FW 1 (sur l'exon 1) et l'amorce RV (sur l'exon 5). C. RT-PCR exécutée avec l'amorce FW 2 (sur l'exon 2) et l'amorce RV (sur l'exon 5). L'ADNc A3B-AS1 est visible à environ 800pb pour le premier couple d'amorces et à environ 400 pour le deuxième couple d'amorces D. RT-qPCR avec les primers FW et RV qPCR A3B-AS1. Les analyses ont été produites sur l'ARN de DND-41, KM-H2 et iPSC avec (+) et sans (-) la rétro-transcriptase. L'amplification spécifique du gène GAPDH a été employée comme contrôle positif de la rétro-transcription.

# Résultats

---

## Caractérisation de A3B-AS1

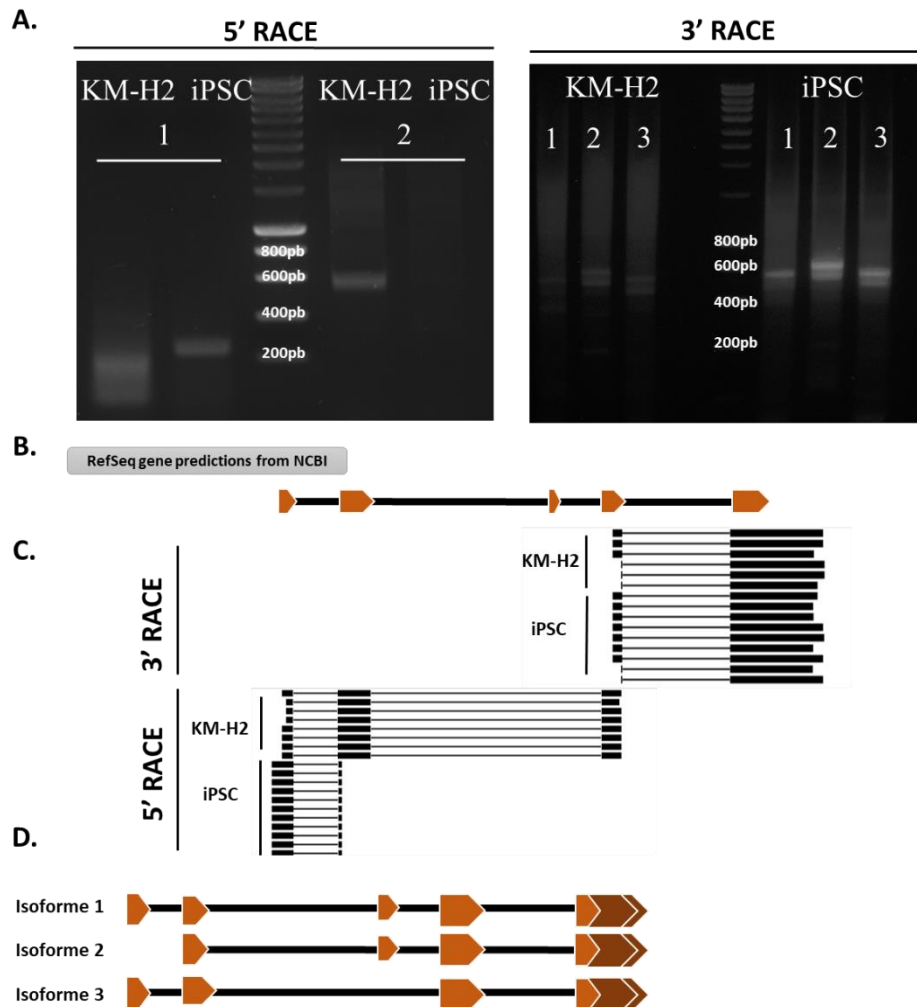
### Mesure de l'expression de A3B-AS1

Bien que l'expression de l'ARNnc A3B-AS1 a été détecté dans des données RNA-seq des lignées DND-41, KM-H2 et IPSC, nous avons dans un premier temps confirmé son expression par analyse par RT-PCR. La première amorce en 5' s'hybride à l'exon 1 et la seconde s'hybride à l'exon 2. L'amorce en 3' s'hybride à l'exon 5 (figure 1.A). Des RT-PCR sans utilisation de la retro-transcriptase ont été réalisées comme contrôle négatif. L'amplification d'ADNc de GAPDH sert quant à elle de contrôle positif ainsi que de contrôle de charge (figure 1.B-C).

Des produits d'amplifications d'environ 800 et 400 paires de bases ont été amplifiés respectivement avec la première et la seconde paire d'amorce dans les lignées cellulaires KM-H2 et IPSC. Ces tailles correspondent aux longueurs attendues d'après la séquence connue du transcrit. L'utilisation des amorces en 5' sur différents exons permet d'assurer l'observation de toutes les isoformes existantes possédant l'exon 5 complémentaire à A3B. Cependant, ces couples d'amorces ne permettent pas d'observer de séquences ne possédant pas les exons 1 et 2 ou de séquences sans l'exon 5 si elles devaient exister.

Par la suite, une analyse de l'expression de A3B-AS1 par RT-qPCR a permis de confirmer les résultats obtenus par RT-PCR. Ces résultats nous permettent également d'estimer que le gène A3B-AS1 est d'avantage exprimé dans les IPSC que dans les KM-H2 bien que cette différence est assez peu marquée (figure 1.D).

Des données de séquençage d'ARN accessibles sur des banques de données en libre accès ont rapporté l'expression d'A3B-AS1 dans deux lignées lymphoïdes (KM-H2 et DND-41) et dans les cellules souches embryonnaires humaines H1. Ces analyses par RT-PCR et RT-qPCR ont confirmé l'expression d'A3B-AS1 dans les KM-H2 ainsi que dans les IPSC mais pas dans la lignée cellulaire DND-41. Il est difficile d'expliquer les raisons pour lesquelles la transcription de A3B-AS1 n'a pas été détectée dans ces cellules. En effet, aucun signal n'a jamais été détecté dans les cellules DND-41 lors de nos analyses RT-PCR et RT-qPCR du gène A3B-AS1. Les cellules KM-H2 apparaissent comme un modèle satisfaisant pour l'étude *in vitro* du gène A3B-AS1.



**Figure 2 : 3' et 5' Race PCR à partir des lignées cellulaires KM-H2 et iPSC.** A. Migration sur gel d'agarose des produits de Race PCR 2 couples de primers ont été testés Pour la RACE en 5' et 3 couples pour la RACE en. B. Représentation du gène A3B-AS1 RefSeq de NCBI et position des primers. Pour la RACE 5', deux RV primers sont utilisés en même temps pour la RT. C. Produits de séquençage mapper sur la séquence prédite de A3B-AS1. D. Grâce au séquençage du produit Race PCR, trois isoformes de A3B-AS1 différentes ont été caractérisées. L'isoforme 1 contient les cinq exons, l'isoforme 2 ne possède pas l'exon 1 et l'isoforme 3 ne contient pas l'exon 3.

## Caractérisation des extrémités 5' et 3' d'A3B-AS1

Nous avons ensuite cherché à caractériser les extrémités 5' et 3' de l'ARN A3B-AS1 par RACE PCR. Les amorces utilisées sont indiquées dans le tableau en annexe 1. Deux amorces FW sont utilisées pour la 5' RACE avec une amorce soit sur l'exon 2 soit sur l'exon 4. Pour les RACE PCR sur l'extrémité 3', ce sont 3 amorces FW différentes qui ont été utilisés. L'analyse des produits de RACE PCR sur gel d'agarose a permis de visualiser les différents fragments obtenus.

Pour la 5' RACE PCR, un produit majoritaire avoisinant la taille attendue d'environ 190 ou 620 pb selon les amorces utilisées d'après la séquence connue est observé pour les iPSC et les KM-

H2. En revanche, la première piste (amorce 5'RV1 s'hybridant à l'exon 2 pour les KM-H2) ne contient qu'une trainée de séquences qui s'est révélé après séquençage être dû à des amplifications non spécifiques. La piste 4 (amorce 5'RV2 s'hybridant à l'exon 4 pour les IPSC) n'a également pas fonctionné.

Pour la 3' RACE PCR, différentes bandes sont observables sur le gel pour chaque combinaison d'amorces et ce pour les deux lignées cellulaires. Le site d'hybridation des amorces FW utilisées pour la 3' RACE PCR chevauche l'exon 1 et 2 pour l'amorce 3' FW1, se situe sur l'exon 2 pour l'amorce 3' FW2 et se situe au début de l'exon 1 pour l'amorce 3'FW3. Les principales bandes observées au niveau de la 3'RACE PCR se situe à environ 600 paires de bases et sont au nombre de 2 pour chaque produit d'amplifications. Ces 2 isoformes majoritaires sont observées dans les produits de séquençage (figure 2.B.C.).

L'analyse par séquençage de l'ensemble des produits PCR a donnée de nombreuses informations sur A3B-AS1 (figure 2.C). Il a été démontré que celui-ci possède des formes différentes dans les 2 lignées cellulaires étudiées. Dans les KM-H2, l'exon 1 débute au niveau de deux sites distincts lui donnant une taille plus courte de 44 et 79 nucléotides par rapport à la séquence de référence de A3B-AS1. L'extrémité 3' possèdent également 2 sites de terminaison au niveau de l'exon 5 lui donnant une taille plus longue de 542 et 620 nucléotides par rapport à la séquence de référence. Il est intéressant de noter que l'on retrouve 4 séquences « AATAA » correspondant à un signal de polyadénylation à la fin de l'exon 5. Une autre information essentielle obtenue grâce à la 5' RACE PCR sur les KM-H2 avec l'amorce 5'RV1 est l'absence d'exon 3 dans cette lignée. Hélas, l'analyse par séquençage du produit de PCR des IPSC avec cette amorce n'a pas permis de retrouver des séquences correspondant à l'A3B-AS1. D'autres résultats de séquençage non reportés ici ont cependant permis d'établir que l'A3B-AS1 possède l'exon 3 dans les IPSC. Concernant les IPSC, l'exon 1 débute bien au même niveau que la séquence de référence. La 5' RACE PCR permet également d'observer qu'une isoforme de A3B-AS1 sans l'exon 1 existe. L'analyse de séquençage de la 3' RACE PCR dans les IPSC montre des résultats identiques à ceux obtenus avec les KM-H2.

Cette analyse par RACE PCR a donc permis de distinguer 3 isoformes principales de A3B-AS1 ainsi que trois extrémités différentes en 5' de l'exon 1 et 2 extrémités différentes en 3' de l'exon 5. L'existence de ces différentes terminaisons au niveau de l'exon 5 pourraient indiquer que plusieurs des séquences « AATAA » retrouvées possèdent un rôle en tant que signal de polyadénylations actif. Il est également intéressant de comparer ces données à A3B qui possède 2 séquences « AATAA » sur son extrémité 3'. La première est complémentaire à l'antisens et la seconde est hors de la séquence de complémentarité.



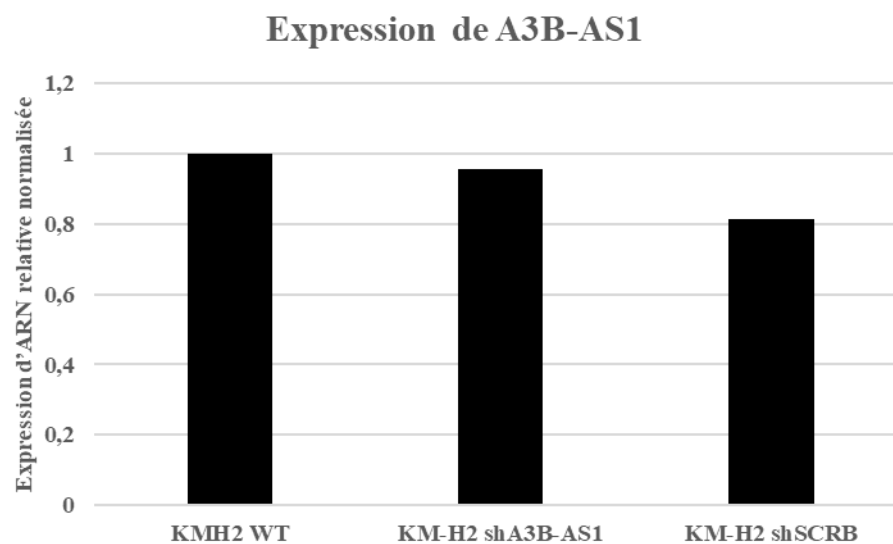


Figure 3 : Analyse par RT-qPCR de l'expression de A3B-AS1 sur les lignées cellulaires KMH2 WT, KMH2 shA3B-AS1 et KMH2 shSCR. Aucune différence significative de l'expression de A3B-AS1 n'est observable entre les trois lignées cellulaires. Les résultats ont été normalisés avec trois gènes de ménages (GAPDH, TBP, HPRT).

## Développement de KM-H2 surexprimant ou sous-exprimant le gène A3B-AS1

Les précédents résultats ayant mis en évidence que la lignée KM-H2 est la seule à exprimer A3B-AS1 avec les IPSC, nous avons décidé d'utiliser cette lignée pour étudier l'impact de la surexpression et sous-expression de notre antisens sur les APOBEC3. L'avantage majeur des KM-H2 par rapport aux IPSC est la facilité de manipulation de ces cellules. La production de lignées stables génétiquement modifiées a donc été faite dans les KM-H2 en priorité pour permettre d'obtenir de premiers résultats rapidement. Cela dit, il est important de noter que le gène A3B-AS1 n'est pas exprimé de la même manière entre ces deux lignées. Il est à prendre en considération que l'ARNlnc A3B-AS1 est plus faiblement exprimé dans les KM-H2 et surtout que les isoformes exprimées sont différentes (isoforme 1 et 2 pour les IPSC et isoforme 3 pour les KM-H2).

### Expression de shRNA contre A3B-AS1 dans les KM-H2

Pour analyser l'effet de la diminution de l'expression de A3B-AS1 dans les KM-H2 sur l'expression transcriptionnelle et traductionnelle des différents APOBEC3. Une lignée de KM-H2 exprimant des shRNA ciblant A3B-AS1 de manière stable a été produite. Comme contrôle, une seconde lignée exprimant un shRNA scramble a été produite. Cette lignée a pour objectif de confirmer que les différences d'expressions de gènes qui auraient potentiellement pu être analysées entre la condition WT et shRNA A3B-AS1 sont bien dues au mécanisme de dégradation du transcrit de l'antisens enclenché par l'action des shRNA. En effet, l'infection par les lentivirus ou l'expression de shRNA dans les KM-H2 pourrait influencer l'expression des APOBEC3 indépendamment de l'expression de l'ARNlnc. Ce qui est d'autant plus vrai en sachant que certains APOBEC peuvent être induits en réponse à une infection virale ou lors de la détection d'ADN ou d'ARN étranger dans le cytoplasme. Les cellules ont été traitées à la blasticidine pour sélectionner celles dont les plasmides contenant les shRNA se sont insérés dans le génome. Ce plasmide exprime le fluorochrome mCherry permettant de contrôler l'efficacité de la sélection.

Pour confirmer l'efficacité des shRNA, une analyse par RT-qPCR du gène A3B-AS1 a été réalisée. Le produit de rétrotranscription a été dilué 10, 30 et 90 fois et chaque dilution a été analysée en duplicats par qPCR. Les résultats ont été normalisés avec trois gènes de ménage (GAPDH, TBP et HPRT). L'histogramme en figure 3 montre l'expression de A3B-AS1 dans les deux lignées modifiées par rapport à la lignée WT. Cette analyse a montré que l'expression des shRNA n'a pas permis de diminuer l'expression de A3B-AS1. En effet, même si de très légères variations sont visibles, aucune différence significative ne peut être observée. L'expression des shRNA n'a donc pas permis de réduire l'expression de A3B-AS1. L'un des shRNA produit cible l'exon 3 de A3B-AS1. Au moment où la transduction des KM-H2 a été réalisée, l'isoforme sans l'exon 3 présente dans les KM-H2 n'avait pas encore été découverte. Seul deux shRNA ciblent alors l'isoforme « complet » de l'antisens. Si seulement le plasmide contenant le shRNA non spécifique de l'isoforme présente dans les KM-H2 a effectivement été transduit, cela pourrait expliquer pourquoi l'expression de A3B-AS1 ne diminue pas alors que

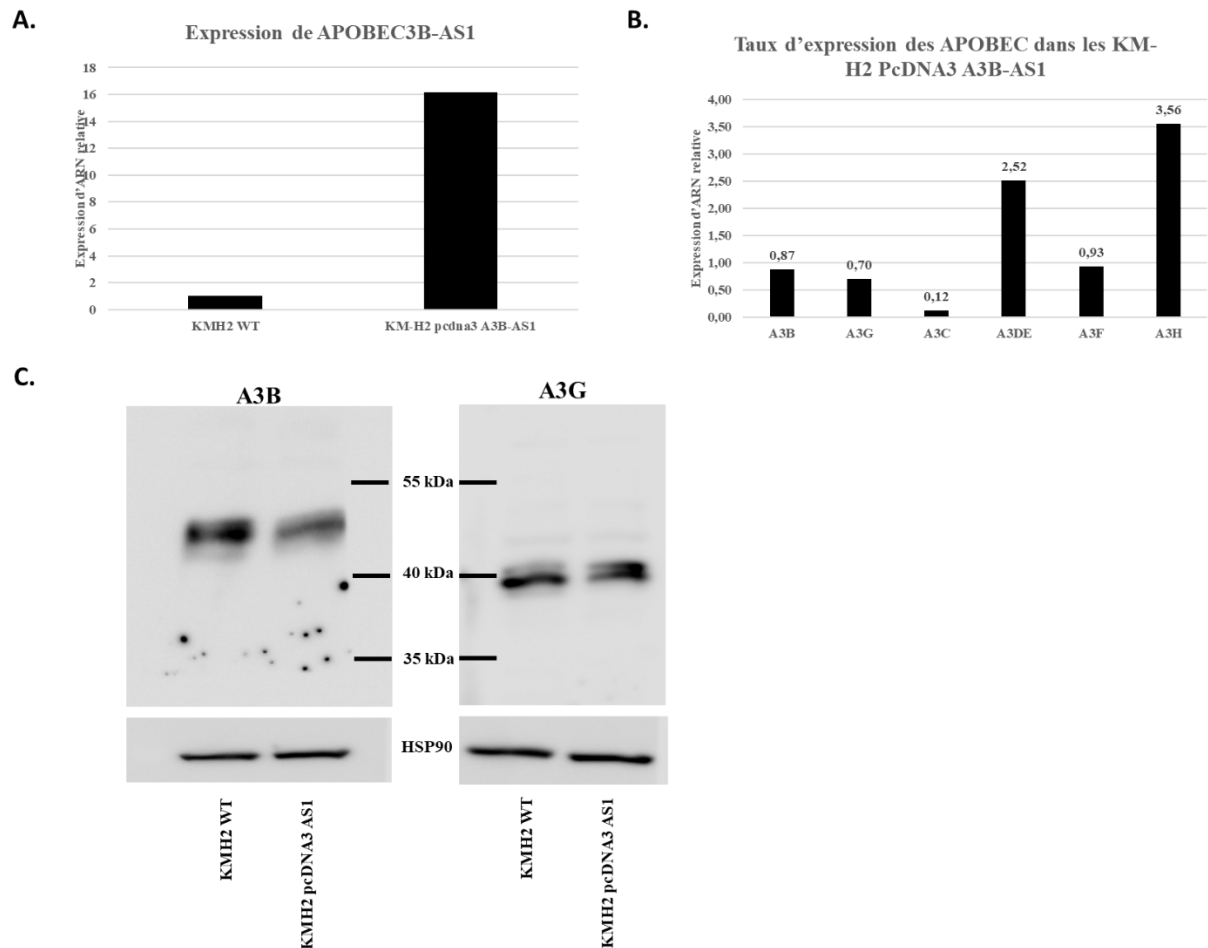


Figure 4 : Analyses par RT-qPCR et par Western-blot de l'expression relative de A3B-AS1 dans les lignées cellulaires KM-H2 WT, KM-H2 PcDNA3 A3B-AS1. A. Comparaison de l'expression transcriptionnelle de A3B-AS1 dans les KM-H2 WT et les KM-H2 pcDNA3.1 (-) A3B-AS1. A3B-AS1 est environ 16 fois plus exprimé dans les KM-H2 nucléofectées avec le plasmide PcDNA3.1 (-) contenant l'isoforme 3 de A3B -AS1. Les résultats ont été normalisés avec trois gènes de ménages (GAPDH, TBP, HPRT). B. Analyse du taux d'expression transcriptionnelle de l'ensemble des gènes APOBEC3 dans les KM-H2 pcDNA3.1 (-) A3B-AS1 par rapport aux KM-H2 WT. L'expression de A3A n'est pas présentée car elle n'a été détectée dans aucune condition. C. Analyse de l'expression protéique de A3B et A3G par Western-blot. Hsp90 est utilisé en tant que contrôle de charge.

l'observation par microscopie en fluorescence montre une expression du mCherry dans toutes les cellules. Cette hypothèse est cependant peu probable car les constructions des différents plasmides codant les shRNA utilisés sont identiques et la méthode utilisée permet normalement l'introduction de plusieurs copies d'entre eux dans chaque cellule.

## **Surexpression de A3B-AS1 dans les KM-H2**

Pour tester l'effet de la surexpression de A3B-AS1 sur l'expression des APOBEC3, une lignée cellulaire de KM-H2 surexprimant l'isoforme 3 de A3B-AS1 a été produite, l'isoforme 1 et 2 étant en cours d'élaboration. Pour cela, les cellules ont été nucléofectées avec le pcDNA3.1 (-) contenant A3B-AS1. Une nucléofection à l'aide du plasmide pMax GFP a également été réalisée pour contrôler que les conditions de transfection étaient correctes par observation de l'expression de la GFP par microscopie à fluorescence. Après sélection à la généticine (G418), l'expression de A3B-AS1 a été mesurée par RT-qPCR pour s'assurer de sa surexpression par rapport aux KM-H2 WT. Cette mesure a permis de montrer que l'expression de l'ARNlnc de A3B-AS1 est environ 16 fois plus importante dans les KM-H2 transfectées que dans les KM-H2 WT (figure 4.A).

Une fois que la surexpression de A3B-AS1 a été confirmée dans les KM-H2, l'expression transcriptionnelle des différents APOBEC3 a été mesurée et comparée entre la condition de surexpression et le WT. Cette analyse révèle qu'aucune différence significative de l'expression des transcrits de A3B, A3G et A3F n'est observée lors de la surexpression de A3B-AS1. Ce qui est plus surprenant, c'est la légère diminution de l'expression de A3C et surtout, l'augmentation de l'expression de A3DE et A3H. Pour rappel, l'ARNlnc A3B-AS1 partiellement complémentaire aux transcrits de A3B, A3A et A3G. Mais aucune complémentarité n'est observable avec les autres APOBEC3. Les différences d'expression observées ne peuvent donc pas être dues à la liaison de A3-AS1 à ces transcrits. Une explication pourrait être que l'activation des gènes A3DE et A3H serait due à la transfection elle-même et à l'introduction d'ADN plasmidique dans les cellules. Aucune explication concernant la diminution de l'expression du gène A3C n'a encore été trouvée. Il faut tout de même noter que malgré les variations d'expression des APOBEC3 dans les deux conditions, l'expression de ces APOBEC3 reste assez faible. En effet, dans les KM-H2, les APOBEC3 les plus exprimés sont A3B et A3C suivis par A3G. L'expression des autres gènes APOBEC3 est très faible voire inexistante pour A3A.

Contrairement à ce que la complémentarité de séquence entre A3B-AS1 et A3B, A3A, et A3G pouvait laisser suggérer, la surexpression de l'antisens n'a pas eu d'effet sur leur quantité de transcrits. D'une manière plus surprenante les niveaux d'ARNm de A3DE et A3H augmentent tandis que celui de A3C est lui diminué de dix fois. Cette analyse n'a donc pas permis de prouver que A3B-AS1 possède un rôle de régulation transcriptionnelle des APOBEC3.

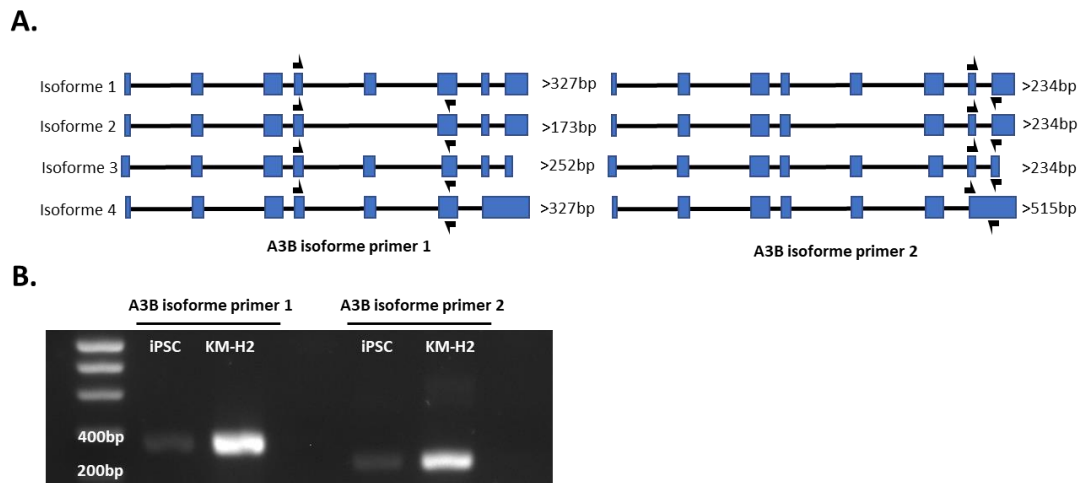


Figure 5 : Analyse des isoformes de APOBEC3B dans les IPSC et KM-H2. A. La première paire d'amorces permet de distinguer l'isoforme 1 et 4 (327pb) de l'isoforme 2 (173pb) et de l'isoforme 3 (252pb). La deuxième paire d'amorces permet de distinguer l'isoforme 1, 2 et 3 (234pb) de l'isoforme 4 (515pb). B. Le produit d'amplification pour la première paire d'amorces a une taille comprise entre 300 et 400 pb pour les 2 lignées cellulaires et la taille du deuxième produit est d'environ 200 pb correspondant à l'isoforme 1.

En complément des niveaux d'expression des transcrits APOBEC3, la présence des protéines a également été analysé par western blot pour A3B et A3G. En effet, même si peu de variation a été détectée au niveau de l'expression des transcrits, A3B-AS1 pourrait posséder une fonction post-transcriptionnelle. Par liaison aux transcrits, l'ARNlnc A3B-AS1 pourrait soit servir de protection contre les nucléases ou les miARN 7 ou 134 et ainsi stabiliser l'ARNm, soit empêcher la liaison du complexe ribosomique permettant la traduction.

L'expression protéiques des A3B et A3G a été mesurée par Western-blot (figure 4.C). L'expression des autres APOBEC3 n'a pas été testée car le laboratoire ne dispose que des anticorps ciblant A3B et A3G. Les protéines A3B et A3G ont bien été détectée dans les KM-H2. Le contrôle de charge utilisé est HSP90. Concernant A3B une légère diminution de l'expression de la protéine est observable dans les KM-H2 surexprimant A3B-AS1. Cependant, aucune variation n'est observée pour A3G. Les résultats obtenus n'ont pas permis de démontrer que A3B-AS1 possède une influence sur l'expression traductionnelle de A3B et A3G. En effet, la variation du niveau protéique de A3B est trop faible pour pouvoir être prise en compte. De plus ce Western-blot n'a été réalisé qu'une seule fois ce qui n'est pas suffisant pour s'assurer de la véracité des résultats. Une seconde bande peut également être distinguée juste au-dessus de A3G. Celle-ci pourrait correspondre à une isoforme de A3G ou une forme phosphorylée mais d'avantage d'investigations sont nécessaires pour le déterminer.

## Étude de la fonction de A3B-AS1

### **Caractérisation des isoformes de A3B dans les IPSC et les KM-H2**

Le gène A3B peut être transcrit en quatre isoformes distinctes. Ces isoformes présentent des extrémités 3' différentes ce qui pourrait avoir un impact sur leurs interactions avec A3B-AS1. Il est donc nécessaire de connaître l'isoforme exprimée dans les lignées cellulaires étudiées. Le but ici est d'identifier l'isoforme de A3B présente dans les KM-H2. Les quatre isoformes du transcrit A3B ont donc été testés par RT-PCR spécifiques dans les lignées cellulaires KM-H2 et IPSC. En se basant sur la composition de la séquence des exons, deux amplifications PCR ont été réalisées afin de discriminer les différentes isoformes d'ADNc. Cette expérience permet de détecter que l'isoforme 1 est le seul présent dans les deux lignées cellulaires (figure 5.A. et 5.B.). Cet isoforme correspond à l'isoforme A3B la plus abondamment retrouvée dans l'ensemble des tissus cellulaires humain. C'est cette isoforme d'A3B qui sera donc utilisée couplée à un gène rapporteur luciférase afin de tester l'effet d'une co-expression A3B et A3B-AS1 sur l'abondance de transcrits A3B.

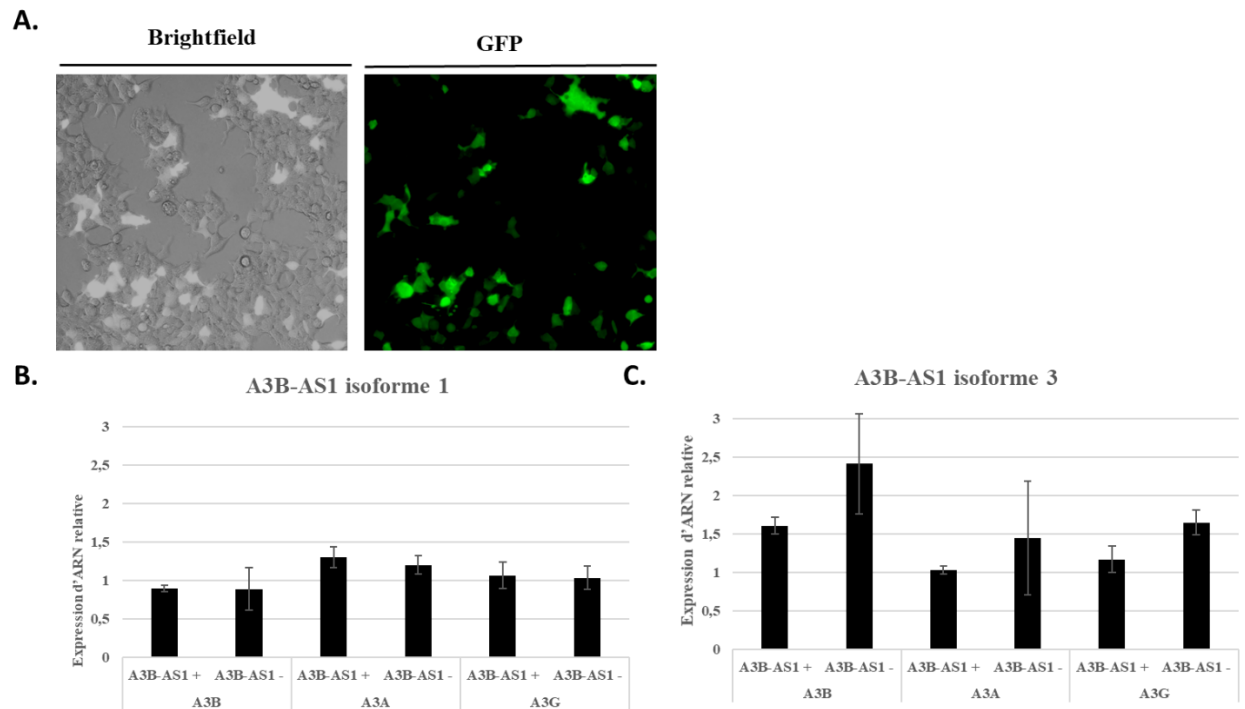


Figure 6 : Analyse de l'influence de l'expression des isoformes 1 et 3 de A3B-AS1 sur l'expression de A3B, A3A et A3G. A. Contrôle de l'efficacité de transfection par analyse du taux d'expression de la GFP dans des HEK293T transfectée avec le pMax GFP. B-C. comparaison des ratios firefly/renilla luciferase pour les différentes conditions étudiées : transfection avec le psi2 contenant la séquence de A3B, A3A ou A3G avec et sans le pcDNA3.1(-) A3B-AS1 isoforme et isoforme 3.

## Impacte de l'expression de A3B-AS1 sur la traduction de A3B, A3A et A3G dans les HEK293T

L'impact de A3B-AS1 sur la traduction de A3B, A3A et A3G a été étudié par mesure de l'activité luciférase couplée avec ces APOBEC. Pour cela, la lignée cellulaire HEK293T a été transfectée par lipofection à l'aide de trois constructions du plasmide psi2. Chacune de ces constructions contient soit A3B, soit A3A soit A3G couplés à un gène codant pour la luciférase. En plus de ces constructions, les cellules ont reçu le plasmide pcDNA3.1(-) contenant A3B-AS1.

Les isoforme 1 et 3 de A3B-AS1 ont été testées. En ce qui concerne A3B, l'isoforme 1 a été utilisée pour la surexpression de manière à correspondre aux conditions retrouvées pour les KM-H2 et les IPSC. Une lipofection avec un plasmide pMax GFP a été réalisée et permet de contrôler l'efficacité de transfection. Seul 25% des HEK293T transfectées par ce plasmide présentent une expression de la GFP par observation en microscopie à fluorescence (figure 6.A). L'efficacité de la transfection est donc relativement faible, ce qui limite l'observation de variations entre les différentes conditions. L'activité de la *firefly* luciférase mesurée par cette méthode est directement proportionnelle au taux de traduction de la construction. En cas d'un effet post transcriptionnelle de l'A3B sur les APOBEC3 on s'attend à observer une variation de la quantité de luciférase firefly disponible. Afin de normaliser les valeurs obtenues par rapport à la quantité d'extrait cellulaire présent dans chaque échantillon, la *Renilla* luciférase est utilisée comme standard interne. Le rapport entre l'activité de la *firefly* et *Renilla* luciférase est ensuite calculé pour les différentes conditions. La manipulation a été faite en plaques 24 puits et chaque condition a été réalisée dans trois puits. Les résultats obtenus montrent très peu de variation de l'activité luciférase entre les cellules qui expriment l'isoforme 1 de A3B-AS1 et le contrôle négatif (figure 6.B). Une légère augmentation de l'activité luciférase est détectée lorsque l'isoforme 3 de A3B-AS1 est exprimée (figure 6.C). Cependant, la dispersion des valeurs obtenues représentées par les barres d'erreurs indique que d'aussi faibles variations ne peuvent être jugées comme étant significatives. De plus, les conditions d'expérimentations n'étaient pas optimales. En effet, l'efficacité de transfection n'est que de 25% environ comme le montre le contrôle GFP. Après 2 jours d'incubation, les cellules n'étaient toujours pas à confluence. Pour améliorer l'efficacité de transfection cette expérience devra être répétée avec un autre kit de lipofectamine (Lipofectamine 3000 Reagent de Thermo Fischer). Les cellules HEK293T sont très sensibles à la toxicité de la lipofectamine ce qui les empêche d'avoir une croissance normale. L'utilisation de ce nouveau kit moins toxique pour les cellules devrait donc pouvoir améliorer l'état des cellules et ainsi leur capacité de transfection et la répétabilité de la manipulation.



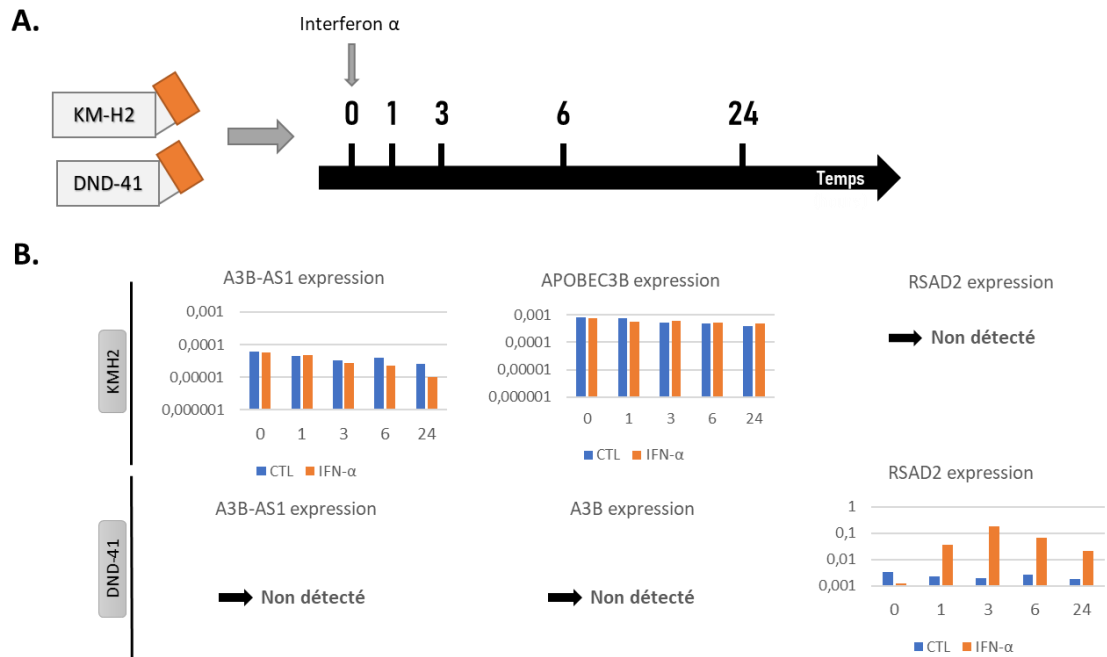


Figure 7 : Effet de l'interféron  $\alpha$  sur l'expression des gènes A3B-AS1, A3B et RSAD2 dans les lignées cellulaires KM-H2 et DND-41. A. Les KM-H2 et les DND-41 ont été incubés avec l'interféron  $\alpha$  (1 000 U/mL) pendant 0, 1, 3, 6 et 24 heures et comparés avec un contrôle non traité. B. l'expression d'A3B-AS1 a été mesurée par qPCR pour savoir si ce gène est interféron inducible. Des analyses qPCR sur les gènes A3B et RSAD2 ont également été réalisées pour déterminer s'il y a une réponse interféron dans les KM-H2 et les DND-41. Les résultats ont été normalisés par le gène GAPDH

## **Impacte de l'interféron $\alpha$ sur l'expression de A3B-AS1 et A3B dans les KM-H2**

Étant donné que l'expression des APOBEC3 peut être modulée par l'interféron  $\alpha$  [6], un possible impact de cette protéine sur l'expression de A3B-AS1 a été étudié. Dans ce but, les lignées cellulaires KM-H2 et DND-41 ont été incubées avec de l'IFN- $\alpha$  pour tester ses effets sur la transcription à la fois de A3B et de l'antisens. Le RSAD2 est un gène inductible par l'IFN et a été utilisé comme contrôle positif. Les extractions d'ARN ont été effectuées après 0, 1, 3, 6 et 24 heures d'incubation (figure 7.A). Dans les cellules KM-H2, l'expression des gènes A3B-AS1 et A3B n'est pas significativement différente entre le contrôle non traité et les cellules traitées avec de l'IFN. L'expression de RSAD2 n'a pas été détectée dans ces cellules. Dans les DND-41, aucune expression du gène A3B-AS1 n'a été détectée. L'expression du gène RSAD2 a fortement augmenté, illustrant la sensibilité du DND-41 à la signalisation de l'IFN- $\alpha$ . Ainsi, le traitement à l'interféron- $\alpha$  n'a pas modifié les niveaux d'expression de A3B-AS1 dans les KM-H2 ou dans les cellules DND-41 (figure 7.B).

L'influence de l'IFN- $\alpha$  sur les gènes A3B et A3B-AS1 n'a donc pas pu être prouvée dans les KM-H2 et les DND-41. En fait, l'absence de détection du gène témoin (RSAD2) dans les KM-H2 empêche d'interpréter correctement l'expression des autres gènes. Les DND-41 et les KM-H2 sont des lignées cellulaires ayant une origine similaire (cellules leucocytaires cancéreuses) et il pourrait être imaginé que leur réponse à l'interféron soit semblable. Cela dit, la détection de réponse à l'interféron dans les KM-H2 reste inconnu car le type de cellule exact de cette lignée est inconnu. En effet, les KM-H2 sont des cellules leucocytaires fortement mutées qui proviennent d'un carcinome thyroïdien anaplastique humain et leurs clusters de différenciation ne correspondent à aucun de ceux des cellules leucocytaires. Comme la réponse à l'interféron est très spécifique au type cellulaire, il n'est pas possible de prédire comment ces cellules réagiraient normalement à l'induction IFN- $\alpha$ .

En conclusion, le traitement l'IFN- $\alpha$  fonctionne sur DND41, mais ne semble pas avoir d'effets sur les KM-H2. Ces cellules ne semblent pas être sensible à l'INF ce qui pourrait indiquer que la présence de A3B-AS1 est indépendante de cette voie. De plus, bien que les DND-41 réagissent à l'IFN- $\alpha$ , aucune induction de l'antisens n'est détectée.



## Discussion

---

Depuis leur découverte, les ARNlnc ont dévoilé de multiples rôles de régulation mais la plupart sont encore méconnus. Les dernières avancées technologiques en analyses génomiques et transcriptomiques ont permis de recenser ces ARN dans différents tissus. De nombreuses études sont en cours pour mieux comprendre ce groupe d'ARN [34], [35]. Dans le cadre de cette étude, nous savons que A3B-AS1 est l'un des gènes les plus exprimés au cours du début de l'embryogénèse mais que cette expression devient quasi inexistante chez l'adulte. Nous pouvons donc émettre l'hypothèse que A3B-AS1 possède un rôle important dans la régulation du développement embryonnaire. Par ailleurs, la régulation de l'embryogénèse par des ARNlnc a déjà été montrée à de multiples reprises. Par exemple, il a été prouvé que l'inhibition de l'expression de l'ARNlnc H19 dans l'embryon provoquait un développement anormal. H19 est le premier ARNlnc à avoir été décrit. Celui-ci mesure 2.3kb et possède une coiffe et une polyadénylation tout comme A3B-AS1 [36]. Il a par la suite été démontré qu'il possède une fonction d'éponge à miARN. Cet ARN est également connu pour être exclusivement exprimé par l'allèle maternel grâce au phénomène d'empreinte parentale [37].

D'autres études comparant des données de RNA-seq à différents stades de développement de l'embryon avec celles de tissus adultes ont permis de montrer que de nombreux ARNlnc sont exprimés très tôt mais également que cette expression est spécifique du stade de développement de l'organisme. En effet, ces études dévoilent qu'une grande partie des ARNlnc exprimé chez l'embryon en phase pré-implantatoire ne sont pas retrouvés chez le fœtus ou l'adulte. De manière encore plus intéressante, certains ARNlnc spécifique du début l'embryogénèse sont retrouvés dans différents cancers. Ce phénomène pourrait être expliqué par une dédifférenciation de certaines cellules cancéreuses vers un état de cellules immatures [38]. La fonction oncogénique de certains de ces ARNlnc a déjà été caractérisée. C'est le cas pour l'antisens ANRIL fréquemment retrouvé dans le cancer de la prostate ayant une fonction d'inhibiteur du suppresseur de tumeur CDKN2B (décrit en introduction). Ceci dit, beaucoup d'ARNlnc sont surexprimés dans des tumeurs sans avoir de fonction oncogénique décrite [39], [40].

Bien que les shRNA se soient révélés inefficaces, il serait très intéressant de déterminer si l'absence de A3B-AS1 ou la diminution de son expression a un impact sur l'expression des APOBEC dans les KM-H2. Pour cela, la transfection à l'aide de lentivirus pourrait être à nouveau effectuée mais cette fois en y intégrant uniquement des shRNA ne ciblant pas l'exon 3 de A3B-AS1. Bien que cela soit peu probable, l'expression de mCherry par les KM-H2 transfectées pourrait être la conséquence d'une efficacité de transfection supérieur pour un seul des plasmides. Il pourrait être judicieux d'utiliser des shRNA complémentaires de chacune des isoformes de A3B-AS1. Une autre hypothèse serait que l'un des shRNA provoque une léthalité des cellules l'exprimant à cause d'un effet non spécifique. Enfin, nous n'avons pas encore d'informations sur la structure secondaire de A3B-AS1. Il serait donc possible qu'une structure secondaire trop complexe de l'antisens notamment au niveau des sites de liaison des shRNA soit la cause de l'inefficacité de cette méthode de sous-expression.

Nous pourrions également nous demander si cette méthode est optimale pour l'étude d'un effet sur les APOBEC. En effet, les APOBEC sont activés lors d'infections virales et sont connus pour provoquer des mutations dans ces virus ainsi que dans le génome de la cellule.



Bien que les lentivirus ne se multiplient pas dans les KM-H2 et sont relativement rapidement éliminer après intégration des shRNA, ceux-ci pourraient activer l'expression des APOBEC indépendamment de la présence de l'antisens. Nous ne connaissons pas l'impact d'une potentielle surexpression même transitoire de ces APOBEC sur les KM-H2 ainsi que sur les séquences intégrées. Pour contrôler l'activation des APOBEC, nous utilisons le contrôle avec un lentivirus intégrant une séquence scramble dans les KM-H2. Si ce modèle avait fonctionné, nous aurions alors pu différencier l'activation des APOBEC par les lentivirus d'une potentielle activation par la surexpression de A3B-AS1. Cette méthode de sous expression sera laissée de côté mais d'autres méthodes pourraient être explorées telles que la mutation ou la suppression de la séquence promotrice de A3B-AS1 au niveau du génome (voir conclusion et perspectives).

Les résultats obtenus après surexpression de A3B, A3A, A3G couplés à une luciférase ainsi que de A3B-AS1 dans les HEK293T n'ont pas conduit à la découverte d'une expression significativement différente. Cependant, le manque d'efficacité de transfection pourrait en être la cause. Ces résultats ne sont donc pas suffisants pour conclure à une absence d'interaction directe entre ces séquences. Il serait intéressant de répéter cette expérimentation en optimisant l'efficacité de la lipofection. Par exemple, la concentration en plasmides pourrait être augmentée ainsi que le temps d'incubation des cellules avec la lipofectamine. En effet, l'efficacité de cette méthode est fortement influencée par le type cellulaire. Nous pourrions également nous questionner sur le fait que dans certaines conditions testées nous souhaitons transfecter plusieurs plasmides alors que dans notre contrôle à l'aide du pMax-GFP un seul plasmide est utilisé. Il est généralement admis que dans la lipofection, l'ajout de 2 plasmides simultanément dans les cellules ne soit pas un facteur limitant. Cela étant dit, si l'efficacité des transfections est faible, certaines cellules peuvent ne recevoir que l'un des plasmides souhaités diminuant ainsi encore les probabilités d'expressions simultanées d'un des APOBEC3 testé avec notre antisens. Pour contrôler l'efficacité de la surexpression de l'antisens A3B-AS1, une RT-PCR aurait pu être réalisée sur une extraction d'ARN après lipofection des cellules avec le plasmide.

La surexpression de A3B-AS1 par nucléofection à l'aide d'un plasmide pcDNA3.1(-) présente l'avantage non négligeable de créer une lignée stable dont l'expression de l'antisens est relativement homogène. Malgré cela, peu de différences sont observables dans l'expression des APOBEC3 complémentaires à l'antisens entre les KM-H2 WT et les KM-H2 surexprimant A3B-AS1. D'importantes différences d'expressions sont en revanche observées pour A3C, A3DE et A3H ce qui était assez inattendu. Il serait intéressant de répéter les RT-qPCR pour permettre une analyse statistique de ces résultats. En effet, bien que l'analyse ait été faite en triplicat technique, aucun réplicat biologique n'a été réalisé ce qui rend ces résultats relativement peu significatifs. Malgré ces résultats peu encourageants, la création de ce plasmide de surexpression ainsi que de cette lignée cellulaire sont à eux seuls des avancées importantes dans l'étude de notre antisens. En effet, d'autres expérimentations pourraient être réalisées dans le futur à partir de ces outils. Il aurait également pu être intéressant de nucléofecter ce plasmide dans d'autres lignées n'exprimant pas l'antisens naturellement. Par exemple, en créant une lignée de HEK293T surexprimant de manière stable A3B-AS1, nous pourrions réaliser de nouveau la lipofection à l'aide des APOBEC3-B -A et -G liés à une luciférase et ainsi s'assurer que l'ensemble des cellules expriment bien l'antisens.



## Conclusion et perspectives

---

Cette étude a permis de rapporter la présence de A3B-AS1 dans les lignées cellulaires KM-H2 et IPSC. Ces deux lignées ont donc été retenues comme modèles cellulaires potentiels dans l'étude de A3B-AS1. Dans cette étude, les analyses ont été réalisées à l'aide des KM-H2 pour permettre d'obtenir des résultats plus rapidement et simplement qu'à l'aide de cellules souches. Les analyses effectuées sur cette lignée n'ont malheureusement pas permis de trouver la fonction de A3B-AS1.

Puisque A3B-AS1 apparaît fortement exprimé dans l'épiblaste au stade E7, un autre très bon modèle d'étude de ce gène serait les cellules souches embryonnaires. L'avantage majeur de ces lignées est leur pertinence biologique car cela permettrait des analyses plus proches des conditions physiologiques que dans les KM-H2 ou même les IPSC. L'inconvénient de ces lignées est la difficulté technique du maintien de leur état de cellules souches. En effet, comme les analyses par RNA-seq le montre, le niveau d'expression de A3B-AS1 est très variant en fonction du stade des cellules dans l'embryogénèse.

Si les hESC sont utilisées comme modèles, l'expression de A3B-AS1 devra être confirmée dans cette lignée cellulaire. Il sera également nécessaire d'analyser quelle(s) isoforme(s) de l'antisens y est/sont exprimé(s). Ce gène a déjà été détecté dans les cellules souches embryonnaires par séquençage d'ARN, mais en quantité moindre que dans l'épiblaste. Par la suite, il sera intéressant de produire des cellules souches embryonnaires A3B-AS1 KO à l'aide de CRISPR Cas9. Les cellules souches embryonnaires exprimant sous certaines conditions Cas9 seront transfectées avec des ARN guides ciblant le promoteur putatif et les premiers exons de la séquence de A3B-AS1. La délétion de A3B-AS1 sera vérifiée par RT-qPCR et 3' RACE PCR et l'impact sur A3B sera évalué aux niveaux transcriptionnels et translationnels à différents stades embryonnaires pour suivre l'évolution de son expression. Il convient de noter que les cellules souches embryonnaires sont considérées comme étant naïves lorsqu'elles correspondent à la masse cellulaire interne de l'embryon préimplantatoire et sont considérées comme étant à l'état primaire lorsqu'elles deviennent des cellules épiblastes post-implantation. Dans l'éventualité où l'expression de A3B-AS1 modifierait l'expression de A3B (ou d'autres A3), l'impact de son expression sur l'activité des rétrotransposons (LINE, SINE, Alu) et des rétrovirus endogènes (ERV) pourrait également être analysé.

Une autre analyse intéressante pourrait être celle de l'impact de A3B-AS1 en tant que potentiel éponge à miARN. En effet, les ARNm de A3B, A3A ou A3G possèdent des sites de liaison aux miR -7 et -134. Ces sites de liaison se retrouvent au niveau de la séquence complémentaire à A3B-AS1. Un des modèles possibles est la stabilisation et la protection de ces transcrits APOBEC3 par l'antisens. Si un effet de l'antisens sur A3B, A3A ou A3G avait été mis en évidence grâce à la méthode de surexpression de ces APOBEC liés à une luciférase, nous aurions pu investiguer l'effet de l'antisens sur ces miARN par cette même méthode. Il suffirait d'ajouter des plasmides de surexpression de ces miARN et d'observer si la surexpression de A3B-AS1 permet de maintenir une expression stable des APOBEC3.

La surexpression de l'antisens pourrait être optimisée par la construction d'un vecteur lentiviral. En effet l'utilisation de Lentivirus pour transduire A3B-AS1 pourrait permettre d'obtenir une surexpression plus élevée et avec un plus haut rendement que par transfection. De plus, les vecteurs lentiviraux possèdent l'avantage de pouvoir être utilisés de manière relativement efficace sur des cellules plus sensibles à faible taux de croissance telles que les IPSC.





Les méthodes de surexpression et de sous-expression précédemment décrites possèdent toutes un défaut commun : elles ne permettent pas de détecter un rôle de régulation co-transcriptionnelle tel que la collision des complexes ARN polymérase décrite dans l'introduction. Pour analyser ce type de mécanismes, il faut pouvoir réguler l'antisens au niveau de sa transcription. Un exemple de méthode pouvant être utilisée est l'inhibition de la séquence promotrice à l'aide de la méthode « CRISP inhibitor ». Cette méthode utilise la même machinerie que CRISPR Cas9 mais la caspase ne possède pas de fonction endonucléase et est fusionnée à la protéine KrAB (Krüppel-associated box). Cette dernière permet d'induire des modifications épigénétiques (méthylation en H3K9) empêchant la liaison d'activateurs de la transcription. En construisant un modèle de CRISP inhibitor capable de lier spécifiquement la séquence promotrice de A3B-AS1 sa transcription pourrait ainsi être inhibée sans modifier la séquence génomique [41]. Ce type d'expérimentation pourrait être dans un premier temps mis au point sur les KM-H2.

La régulation de la séquence promotrice de A3B-AS1 pourrait également être étudiée. Pour ce faire, cette séquence promotrice pourrait être clonée dans un plasmide de manière à activer l'expression d'une luciférase. Ce plasmide pourrait ainsi être intégré dans différentes lignées cellulaires. L'expression de la luciférase pourrait par la suite être analysée en présence ou non de différents facteurs de transcription. Il a récemment été observé que l'activation de NF- $\kappa$ B induit l'expression de A3B dans certains cancers du col de l'utérus causés par HPV (Human Papillomavirus) [42]. Une potentielle activation de l'antisens par NF- $\kappa$ B pourrait alors être étudiée par cette méthode. L'un des avantages de l'étude de l'activité de la luciférase par rapport à l'étude de l'expression du transcrit A3B-AS1 est la haute sensibilité de la méthode. De plus, il pourrait être intéressant d'exprimer ce plasmide dans une lignée exprimant l'antisens (KM-H2) ainsi que dans une lignée ne l'exprimant pas (HEK293T). De cette manière, la régulation de la séquence promotrice pourrait être analysée lorsqu'elle est supposée être active ou inactive selon la lignée.

Jusque maintenant, les seules lignées cellulaires cancéreuses dans lesquelles l'expression de A3B-AS1 a été prédit par RNA-seq sont les KM-H2 et DND-41. Il a été démontré dans ce travail par RT-PCR et par RT-qPCR que les DND-41 n'expriment finalement pas l'antisens. De nombreuses lignées cancéreuses n'ont cependant pas encore été analysées. Il serait intéressant d'étudier l'expression de l'antisens dans d'autres lignées cellulaires cancéreuses. Si A3B-AS1 était exprimé dans plusieurs lignées cancéreuses relativement similaires, nous pourrions étudier une potentielle corrélation entre son taux d'expression et celui de A3B.

Pour l'instant, seuls les effets de A3B-AS1 sur les APOBEC ont été étudiés. Il pourrait être intéressant d'analyser si cet ARNlnc antisens possède une fonction de régulation sur d'autres éléments. Pour cela une méthode d'analyse haut débit de l'expression transcriptionnelle telle que du RNA-seq pourrait être mise au point à partir de lignées exprimant ou surexprimant A3B-AS1.

La caractérisation de cet ARNlnc a permis de mettre en évidence 3 isoformes ainsi que différents sites d'initiation et de terminaison de la transcription. Le traitement à l'IFN- $\alpha$  n'a pas modifié l'expression de A3B-AS1. La surexpression et la sous-expression de l'ARNlnc dans KM-H2 ont permis d'étudier plus en détail le rôle de A3B-AS1.



Ce travail aura permis de créer des données préliminaires dans l'étude de la fonction de A3B-AS1 en permettant sa caractérisation. Bien que les données obtenues n'aient pas abouti à définir la fonction exacte de cet antisens, les résultats nous ont permis de mieux définir les objectifs de recherche. En effet, nous savons désormais que la présence d'interféron-  $\alpha$  n'intervient pas dans l'expression de A3B-AS1. Il n'est alors pas utile de continuer à investiguer dans cette voie. De plus, même si les résultats obtenus avec les shRNA n'ont pas été concluants, nous savons désormais que nous pouvons écarter, du moins temporairement, ce type d'expérimentations pour se consacrer à d'autres méthodes.



## Bibliographie

- [1] M. Ohainle, J. A. Kerns, H. S. Malik, and M. Emerman, "Adaptive Evolution and Antiviral Activity of the Conserved Mammalian Cytidine Deaminase APOBEC3H," vol. 80, no. 8, pp. 3853–3862, 2006.
- [2] S. L. Sawyer, M. Emerman, and H. S. Malik, "Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G," *PLoS Biol.*, vol. 2, no. 9, pp. 1278–1285, 2004.
- [3] I. B. Rogozin *et al.*, "Evolution and diversification of lamprey antigen receptors : evidence for involvement of an AID-APOBEC family cytosine deaminase," vol. 8, no. 6, pp. 647–656, 2007.
- [4] L. B. Alexandrov and M. R. Stratton, "Mutational signatures : the patterns of somatic mutations hidden in cancer genomes," *ScienceDirect*, vol. 24, pp. 52–60, 2014.
- [5] J. D. Salter, R. P. Bennett, and H. C. Smith, "The APOBEC Protein Family: United by Structure, Divergent in Function," *Trends Biochem Sci*, vol. 41, no. 7, pp. 578–594, 2017.
- [6] D. A. Covino, "Understanding the regulation of APOBEC3 expression : Current evidence and much to learn," *J. Leukoc. Biol.*, vol. 103, no. 3, pp. 433–444, 2017.
- [7] K. Gillick, D. Pollpeter, P. Phalora, E. Kim, S. M. Wolinsky, and H. Malim, "Suppression of HIV-1 Infection by APOBEC3 Proteins in Primary Human CD4  $\gamma$  T Cells Is Associated with Inhibition of Processive Reverse Transcription as Well as Excessive Cytidine Deamination," vol. 87, no. 3, pp. 1508–1517, 2013.
- [8] S. Stavrou and S. R. Ross, "APOBEC3 Proteins in Viral Immunity," *J. Immunol.*, vol. 195, no. 10, pp. 4565–4570, 2015.
- [9] S. Petropoulos, D. Edsga, B. Reinius, and S. Linnarsson, "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Resource Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos," *Cell*, vol. 165, no. 4, pp. 1012–1026, 2016.
- [10] V. Pelechano and L. M. Steinmetz, "Gene regulation by antisense transcription," *Nat. Publ. Gr.*, vol. 14, no. 12, pp. 880–893, 2013.
- [11] J. J. Quinn and H. Y. Chang, "Unique features of long non-coding RNA biogenesis and function," *Nat. Publ. Gr.*, vol. 17, no. 1, pp. 47–62, 2016.
- [12] C. M. Okeoma, N. Lovsin, B. M. Peterlin, and S. R. Ross, "APOBEC3 inhibits mouse mammary tumour virus replication in vivo," *Nature*, vol. 445, no. February, pp. 927–930, 2007.
- [13] A. M. Sheehy, N. C. Gaddis, J. D. Choi, and M. H. Malim, "Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein," *Nature*, vol. 418, no. August, pp. 4–8, 2002.



- [14] T. K. Naruse *et al.*, "APOBEC3H polymorphisms and susceptibility to HIV-1 infection in an Indian population," *Nat. Publ. Gr.*, no. April, pp. 1–3, 2015.
- [15] P. Jern, R. A. Russell, V. K. Pathak, and J. M. Coffin, "Likely Role of APOBEC3G-Mediated G-to-A Mutations in HIV-1 Evolution and Drug Resistance," *PLoS Pathog.*, vol. 5, no. 4, 2009.
- [16] A. L. Macmillan, R. M. Kohli, and S. R. Ross, "APOBEC3 Inhibition of Mouse Mammary Tumor Virus Infection : the Role of Cytidine Deamination versus Inhibition of Reverse," *J. Virol.*, vol. 87, no. 9, pp. 4808–4817, 2013.
- [17] S. Sanchez-martinez *et al.*, "Studies on the Restriction of Murine Leukemia Viruses by Mouse APOBEC3," vol. 7, no. 5, pp. 1–7, 2012.
- [18] P. J. Thompson, T. S. Macfarlan, and M. C. Lorincz, "Review Long Terminal Repeats : From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire," *Mol. Cell*, vol. 62, no. 5, pp. 766–776, 2016.
- [19] J. Göke and H. H. Ng, "CTRL + INSERT : retrotransposons and their contribution to regulation and innovation of the transcriptome," *EMBO Rep.*, vol. 15, pp. 1–14, 2016.
- [20] Y. Yin, L. Zhou, and S. Yuan, "Enigma of Retrotransposon Biology in Mammalian Early Embryos and Embryonic Stem Cells," *Stem Cells Int. histone*, vol. 2018, pp. 1–6, 2018.
- [21] E. J. Grow *et al.*, "Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells," *Nature*, vol. 522, no. 7555, pp. 221–246, 2015.
- [22] L. B. Alexandrov and M. R. Stratton, "Mutational signatures: The patterns of somatic mutations hidden in cancer genomes," *Curr. Opin. Genet. Dev.*, vol. 24, no. 1, pp. 52–60, 2014.
- [23] M. E. Olson, R. S. Harris, and D. A. Harki, "APOBEC Enzymes as Targets for Virus and Cancer Therapy," *Cell Chem. Biol.*, vol. 25, no. 1, pp. 36–49, 2018.
- [24] E. K. Law *et al.*, "The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer," *Sci. advances*, vol. 2, no. 10, pp. 1–10, 2016.
- [25] E. Malek, S. Jagannathan, and J. J. Driscoll, "Correlation of long non-coding RNA expression with metastasis, drug resistance and clinical outcome in cancer," *Oncotarget*, vol. 5, no. 18, pp. 8027–8038, 2014.
- [26] H. Qu and X. Fang, "A Brief Review on the Human Encyclopedia of DNA Elements (ENCODE) Project," *Genomics, Proteomics Bioinforma.*, vol. 11, no. 3, pp. 135–141, 2013.
- [27] J. Jarroux, A. Morillon, and M. Pinskaya, *History , Discovery , and Classification of lncRNAs*. 2017.
- [28] M. B. Clark *et al.*, "Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing," *Nature*, vol. 522, no. 4, pp. 339–342, 2015.





- [29] V. Pelechano and L. M. Steinmetz, "Gene regulation by antisense transcription," *Nat. Rev. Genet.*, vol. 14, no. 12, pp. 880–893, 2013.
- [30] R. J. Taft, M. Pheasant, and J. S. Mattick, "The relationship between non-protein-coding DNA and eukaryotic complexity," pp. 288–299, 2007.
- [31] F. Rashid, A. Shah, and G. Shan, "Long Non-coding RNAs in the Cytoplasm," *Genomics, Proteomics Bioinforma.*, vol. 14, no. 2, pp. 73–80, 2016.
- [32] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "insights into functions," *Nature*, vol. 10, no. 3, pp. 155–159, 2009.
- [33] B. Gelfand *et al.*, "Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast," *Mol. Cell. Biol.*, vol. 31, no. 8, pp. 1701–1709, 2011.
- [34] L. Yan *et al.*, "Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells," *Nat. Publ. Gr.*, vol. 20, no. 9, pp. 1131–1139, 2013.
- [35] G. Yang, X. Lu, and L. Yuan, "LncRNA: A link between RNA and cancer," *BBA - Gene Regul. Mech.*, vol. 20, no. 9, pp. 1131–1139, 2014.
- [36] M. A. Ripoché, C. Kress, F. Poirier, and L. Dandolo, "Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element," *Genes Dev.*, vol. 11, no. 12, pp. 1596–1604, 1997.
- [37] P. Monnier, C. Martinet, J. Pontis, I. Stancheva, S. Ait-Si-Ali, and L. Dandolo, "H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 51, pp. 20693–20698, 2013.
- [38] J. Bouckenheimer *et al.*, "Long non-coding RNAs in human early embryonic development and their potential in ART," *Hum. Reprod. Update*, vol. 23, no. 1, pp. 19–40, 2016.
- [39] A. Sanchez Calle, Y. Kawamura, Y. Yamamoto, F. Takeshita, and T. Ochiya, "Emerging roles of long non-coding RNA in cancer," *Cancer Sci.*, vol. 109, no. 7, pp. 2093–2100, 2018.
- [40] N. Hauptman and D. Glavač, "Long non-coding RNA in cancer," *Int. J. Mol. Sci.*, vol. 14, no. 3, pp. 4655–4669, 2013.
- [41] P. I. Thakore *et al.*, "Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements," *Nat. Methods*, vol. 12, no. 12, pp. 1143–1149, 2015.
- [42] S. Tilborghs *et al.*, "The role of Nuclear Factor-kappa B signaling in human cervical cancer," *Crit. Rev. Oncol. Hematol.*, vol. 120, no. November, pp. 141–150, 2017.



## Annexes

Tableau 1: Liste des amorces.

Name	Sequence
A3B-AS1 FW primer 1	TCGAGAAGCTTGGATCCTACCAACATGTGATG
A3B-AS1 FW primer 2	TCGAGGAATTCCCTATATAAGGAGGCGCTGCA
A3B-AS1 RV primer	TCGAGAAGCTTAGGGAGACTGGCTGCTGGCTG
5' RACE RT primer 1	CTGCTGATGCAGAAGTCCAG
5' RACE RT primer 2	GCTCACTATGGGTCCTCGTT
GeneRacer™ 5' FW primer	CGACTGGAGCACGAGGACACTGA
5' race A3B-AS1 RV primer 1	GGCCCTGAGTGGGTGCTCTT
5' race A3B-AS1 RV primer 2	GCATCCCAGGTTCTGCCTGA
GeneRacer™ 5' Nested FW Primer	GGACACTGACATGGACTGAAGGAGTA
5' race A3B-AS1 nested RV primer	AAGTGTCGGCTGGCTTGAGA
GeneRacer™ 3' RV primer	GCTGTCAACGATACGCTACGTAACG
3' race A3B-AS1 FW primer 1	AGAGACTGAGGCCCATCCTTCA
3' race A3B-AS1 FW primer 2	CGGTCTGTTTGCATTTCTTGGA
GeneRacer™ 3' Nested RV Primer	CGCTACGTAACGGCATGACAGTG
3' race A3B-AS1 nested FW primer	AGCAGGAGGTCATGATGGAGACTT
qPCR A3B-AS1 FW primer	AGCCAGCCGACACTTAAGAA
qPCR A3B-AS1 RV primer	GCCCACAAGTGGATAGAGGA
qPCR A3B FW primer	GACCCTTTGGTCCTTCGAC
qPCR A3B RV primer	GCACAGCCCCAGGAGAAG
qPCR A3A FW primer	GAGAAGGGACAAGCACATGG
qPCR A3A RV primer	TGGATCCATCAAGTGTCTGG
qPCR A3G FW primer	CCGAGGACCCGAAGGTTAC
qPCR A3G RV primer	TCCAACAGTGCTGAAATTCTG
qPCR RSAD2 FW primer	CAAGACCGGGGAGAATACCTG
qPCR RSAD2 RV primer	AACTCTACTTTGCAGAACCTCAC
qPCR GAPDH FW primer	AATCCCATCACCATCTTCCAG
qPCR GAPDH RV primer	CAGAGATGATGACCCTTTTGG
qPCR HPRT FW primer	GGTCAGGCAGTATAATCCAAAG
qPCR HPRT RV primer	AAGGGCATATCCTACAACAAAC
qPCR TBP FW primer	ACCTAAAGACCATTGCACTTCG
qPCR TBP RV primer	CATATTTTCTTGCTGCCAGTCTG
A3B isoform FW primer 1	ATGCCTTGGTACAAATTCGATG
A3B isoform RV primer 1	CCAAGTGACCCTGTAGATCTGG
A3B isoform FW primer 2	CTTTGTGTACCGCCAGGGAT
A3B isoform RV primer 2	TTTGCTGGTGTCTGTGAGCA